UNIVERSITY OF
LIVERPOOL

Development of a chick model to identify *LSAMP* gene expression enhancers.

Thesis Submitted in accordance with the requirements of the

University *of* Liverpool

for the degree of Doctor of Philosophy

By

Michael Lyons

May 2012

DECLARATION

DECLARATION

The research work was carried out in the Development Neurobiology Research Group, led by Dr. Diana Moss, in the Human Anatomy and Cell Biology Division of the School of Biomedical Sciences, University of Liverpool.

ABSTRACT

One of the major ways employed by eukaryotic cells to control gene expression is by the action of the regulatory sequences known as enhancers and repressors. An understanding of the spatial and temporal activation of these sequences would help our understanding of gene control. Because polymorphisms in these control regions have been associated with disease a way to rapidly identify if these polymorphisms were affecting gene expression would be a useful diagnostic tool.

This thesis describes the development of a chick-based model to attempt to rapidly identify when and where enhancers are being activated. Using the ECR Browser, putative enhancer sequences are identified by cross-species comparison of the genomes to expose evolutionary conserved regions. By cloning an ECR, so identified, into a suitable reporter vector and transfecting cells *in-ovo,* the activation of the enhancer by the cell can be detected. Following a successful proof-of-concept run using the *Hb9* gene, *LSAMP,* was selected as the first gene to investigate because of its role in development and its implication in mental health disorders and cancer. A number of ECRs were examined and six of these were shown to be enhancers proving that the system functioned.

DEDICATION

For Eleanor

ACKNOWLEDGEMENTS

I am very grateful to Diana Moss for the opportunity and immense support to undertake this project and thesis and for her excellent supervision throughout this course.

Great support was also provided by Christine Jane McNamee, who supplied help, friendship and lifts home.

I would also like to thank Christine Cashman for her help, even if our taste in music did not always coincide.

Support and assistance also came from departmental staff, other members of the group, both past and present, Bob Reid, Mohammed Akeel and Rachel Carter.

Thanks too for the reporter plasmid that was used in this project being supplied through John Quinn from Dr. Alasdair MacKenzie, University of Aberdeen.

ABBREVIATIONS

Abbreviation	Expansion of Term
AMP	Ampicillin
AMP^r	Ampicillin resistant
BLAST	Basic local alignment search tool
BLAT	BLAST-like alignment tool
Bp/bp	base pair/base pairs
bpm	Beats per minute
CEPU-1	Cerebellum-Purkinje (chick homolog of NTM)
DAPI	4,6-diamidino-2-phenylindol
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides
DSM	Diagnostic and Statistical Manual of Mental Disorders
dT	Deoxythymidinetriphosphate
E	Embryonic day, number of days following start of incubation.
EBI	European Bioinformatics Institute
ECM	Extra-cellular matrix

ECR	Evolutionary Conserved Region
ECRx:p1229	*LSAMP* enhancer sequence ECRx (where x = 1,2 or 3) and p1229 plasmid construct
EDTA	Ethylenediaminetetraacetic acid
GFP	Green fluorescent protein
GTF	General transcription factor
GWAS	Genome-wide association study
GxEy:p1229	*LSAMP* enhancer sequence GxEy (where x = group number and y = ECR number within that group) and p1229 plasmid construct
Hb9	Homeo box-HB9 (Note: Official name MNX1)
Hb9:p1229	Hb9 enhancer sequence and p1229 plasmid construct
Hox	Homeobox
Inr	Initiator
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRES	Internal ribosome entry site
KAN	Kanamycin
KAN^r	Kanamycin resistant
Kb	Kilo bases
LAMP	Protein encoded for by the *LSAMP* gene.
LB	Lysogeny broth
LSAMP	Limbic system-associated membrane protein gene
M	Molar
MCS	Multiple Cloning Site
mg	Milligram

ml	millilitre
mM	millimolar
MNX1	Motor neuron and pancreas homeobox 1 (a/k/a Hb9)
mRNA	Messenger RNA
NaCl	Sodium Chloride
ng	nanogram
nm	Nanomolar
NTM	Neurotrimin
O.D.	Optical Density
OBCAM	Opiate-binding cell adhesion molecule
OCT medium	Optimal Cutting Temperature medium Bright Cryo-M-Bed medium
p1229	A non-commercial reporter vector based on a pBluescript backbone.
PCR	Polymerase chain-reaction
pCRII	A propriety plasmid used in Invitrogen's TOPO TA cloning kit
PI	Post-injection
PVC	Polyvinyl chloride
RCAS	Replication-Competent ASLV long terminal repeat (LTR) with a Splice acceptor
RE	Restriction enzyme
SAP	Shrimp alkaline phosphatase
SOC	Super Optimal Broth
STET	Sodium, Tris-Cl,EDTA,TRITON X100

SyG	SYBR Green I nucleic acid stain
T_A	Annealing temperature
TAE	Tris acetate EDTA buffer
Tris (a pH buffer)	Tris(hydroxymethyl)aminomethane
TSS	Transcription Start Site
UTR	Untranslated region
V	volts
X-gal	bromo-chloro-indolyl-galactopyranoside
λ	Wave length
µg	microgram
µl	microlitre

TABLE OF CONTENTS

Contents

Chapter 1 - Introduction

1.1 Transcription initiation is an important point at which gene expression is regulated.

The central dogma of molecular biology as defined by Crick was that DNA makes RNA makes Protein (Crick, 1958, Crick, 1970). This process whereby the genetic information held in the form of DNA ultimately makes protein must first go through the transcription of the DNA into RNA. Because each cell in an organism has the same complement of genes the control of what genes are transcribed at a given time in a cell must be under exquisite control. The transcription step is probably the most important control point in this process because it is the most efficient place to moderate the process. In eukaryotic cells the transcription that eventually results in protein and peptides is performed by RNA Polymerase II (PolII). This requires that a plethora of basal transcription machinery assembles on the promoter region upstream of the gene's transcription start site (TSS) in what is known as the pre-initiation complex (PIC). Once this PIC has formed transcription can occur at a basal level but the rate of transcription starts is greatly increased by the receipt of a signal by DNA. However because the DNA is held in the form of chromatin, i.e. DNA wrapped around histone cores, access to the nucleotides is not always possible and the histone core needs to be displaced or its grip on the nucleotides needs to be loosened in order for the signal to be effective. Access to the DNA is implemented by a number of stimuli but mainly by molecules known as Transcription Factors (TFs). These TFs bind to regions of DNA known as 'enhancers'. Enhancers are regions of DNA to which multiple TFs bind and increase the transcription start rate (Banerji et al., 1981). The mechanism behind this seems to be that the transcription factors on the enhancer sequences recruit enzymes such as histone acetyltransferase (HAT) to the targeted DNA domain. HATs acetylate lysine residues in the N-terminal tail function as part of gene regulation (Sadoul et al., 2008) possibly providing binding sites for specific protein–protein interaction domains (Dhalluin et al., 1999). Certain transcription factors

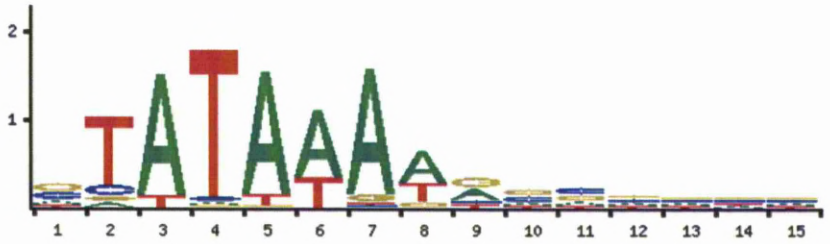
interact with those at the promoter (Maston et al., 2006, Malik and Roeder, 2010) resulting in the formation of a chromatin loop that physically brings together the enhancer and promoter sequences (Schoenfelder et al., 2010, Dekker et al., 2002). Conversely repressor sequences contain TFBS that recruit enzymes such as histone de-acetylase (HDAC) that have the opposite effect.

1.2 Transcription Factor Binding Sites are short regions of DNA to which Transcription Factors bind.

The Transcription Factor Binding Sites (TFBS), to which the TFs bind, are typically in the range of 6-12 bp although the actual binding specificity is typically conferred by 4-6 of these bp. The TFBSs for any given TF are usually degenerate (Zhang et al., 2006) and defined by a consensus sequence in which some locations are more highly conserved than others. Therefore a particular TF can bind to sites that contain differing nucleotides although there would typically be a relatively well conserved central motif. Examples of degenerate binding sequences for two common TFs, TATA Binding Protein (TBP) and murine GATA1 are shown below (Figure 1) showing pictorially and by frequency matrix the relative occurrence of the nucleotides in the binding sites.

Heterodimers and homodimers are often formed by the TFs resulting in the binding sites consisting of two half sites.

A1



A2

A	[61	16	352	3	354	268	360	222	155	56	83	82	82	68	77]
C	[145	46	0	10	0	0	3	2	44	135	147	127	118	107	101]
G	[152	18	2	2	5	0	20	44	157	150	128	128	128	139	140]
T	[31	309	35	374	30	121	6	121	33	48	31	52	61	75	71]

B1

B2

A	[1423	708	2782	0	4000	27	3887	3550	799	1432	1487]
C	[560	1633	31	0	0	29	0	4	681	897	829]
G	[1242	1235	10	4000	0	109	6	383	2296	1360	1099]
T	[775	424	1177	0	0	3835	107	63	224	311	585]

Figure 1 Transcription factor Binding Sites of (TATA Binding Protein) TBP and mouse GATA1.

(A1) Pictogram of the TBP TFBS motif. The letters on the top of each position represent the TBP consensus. The height of each position reflects the information content at that position with the relative height of a letter in each position reflects the frequency of observing the nucleotide.

(A2) Frequency Matrix of TBP underlying model showing the DNA pattern. The cell numbers indicate the number of sequences having base x in column y.

(B1) Pictogram of the GATA1 TFBS motif in mouse. The letters on the top of each position represent the GATA1 consensus. The height of each position reflects the information content at that position with the relative height of a letter in each position reflects the frequency of observing the nucleotide.

(A2) Frequency Matrix of GATA1 underlying model showing the DNA pattern. The cell numbers indicate the number of sequences having base x in column y.

1.3 Identifying DNA regulatory sequences is difficult

Identifying enhancers or repressors by analysing DNA sequences is difficult. This is because of their small size and degeneracy, they have few identifiable motifs, they may be located upstream, downstream or within the transcription unit, can function in either orientation, and may be tens to hundreds of kilobases away from the gene that they control (Banerji et al., 1981) , and may even reside on a different chromosomes (Lomvardas et al., 2006, Geyer et al., 1990). For example, an enhancer of tyrosine kinase was located at a distance of 212 kbp from the gene (Haupt et al., 2010), one for *Myf5* was found between 50 kbp and 95 kbp upstream (Hadchouel et al., 2000)and one for human β -globin was identified at 260 kbp from the gene's TSS (Porcu et al., 1997). One method of identifying enhancers is by finding the binding sites for the p300 protein. The p300 protein interacts with numerous TFs and increases the expression of their target genes (Vo and Goodman, 2001, Kasper et al., 2006). The ChIP-Seq method can be used to analyse the binding of the P300 protein to DNA and help identify enhancers (Visel et al., 2009).

1.4 Many non-coding Evolutionary Conserved Regions function as transcriptional control regions.

A method of identifying putative *cis* regulatory elements is by locating regions of non-coding DNA that are conserved over evolution across many species. These are known as Evolutionary Conserved Regions (ECRs). It was presumed that as these regions are conserved across evolutionary time they must have a function and it is now generally accepted that one of the functions of ECRs is to act as enhancers and repressors of gene expression (Pennacchio et al., 2006, Woolfe et al., 2005). Finding ECRs has been made much easier by the introduction of the databases of such regions and tools to interrogate them such as at University of California at Santa Cruz (Kent et al., 2002, Lindblad-Toh et al., 2011, Pennacchio et al., 2006, Woolfe et al., 2005) and the ECR Browser (Ovcharenko et al., 2004).

The ECR Browser has been previously used in a number of projects such as investigation into distal regulatory elements of *Fshr* expression (Hermann et al., 2007), comparative study of the insulin gene promoters (Hay and Docherty, 2006) and continues to be employed, for example into mouse *Elf5* regulation (Pearton et al., 2011). The ECR Browser is a tool for visualising evolutionary relationships between genomes (Ovcharenko et al., 2004) and was used in this thesis. In this tool one organism is set as the 'base' and the other selected organisms, the 'targets', are compared against this.

The comparison criteria used to identify such ultra-conserved regions typically consists of length of sequence and percentage similarity between the sequence of the base organism and that of the target(s). What is reported as being an ECR will vary depending on how these parameters are set, if set too stringently ECRs will be missed and if the parameters are not sufficiently stringent there will be too many false positives.

The evolutionary distance between the organisms being compared greatly affects the results. If there is too great a difference between the two sequences being compared few or no ECRs will be identified, conversely if there is too small a difference between the organisms too many regions will tend to be flagged as ECRs.

Having identified a region of DNA as being a potential enhancer it must be tested for function. Although great strides have recently been made in deciphering genomes and bioinformatics-based methods of interpreting the instructions encoded in the genome, and in identifying variations associated with disease have improved, they still need to be proven in a living model. This model may be *in vivo* or *in vitro*. *In vivo* is preferable as there is minimal disruption to the cells being examined and the cells can develop within the whole animal and so be subjected to the normal range of growth factors and signals from many cells and the general developmental environment.

1.5 Model Organisms are non-human species that are studied to understand biological functions.

When studying human development and disorders probably the best organism to use would be human however due to a number of reasons, such as ethics, complexity of humans and development time this is not typically practicable. To ameliorate some of these problems model organisms are used. These are typically more simple organisms than the one being studied and their selection and suitability is based on a number of criteria, for example how long they take to regenerate, fecundity, the accessibility to tissues, organs and/or DNA, ease of manipulation, cost, existing knowledge of the organism and similarity to the target organism. Many organisms are used, from the simplest organisms such as yeast and bacteria through *D. melanogaster* and on to the more complex animals such as fish, avians, rodents and primates. Evolution is regarded as a process where populations are altered over time and may divide into separate branches, join together or become extinct. Thanks to the development of phylogeny (Woese et al., 1990, Doolittle, 1999) , the study of evolutionary relationship between organisms through molecular and morphological methods, the evolutionary distance between the entire or even part of an organism can be calculated. This may be visualized in as a phylogenetic tree, a graphic representation of the order in which evolutionary events are assumed to have occurred. Figure 2 indicates the evolutionary distance between the isoforms of *LSAMP* in the selected organisms with the length of the lines indicating the evolutionary distance. As can be seen, of the organisms included in the

comparison, Zebrafish is the most evolutionary distant from human and Pig is the closest.

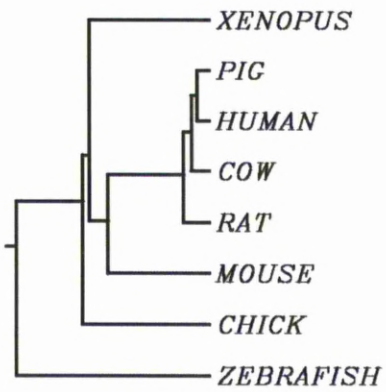


Figure 2. Phylogenetic tree showing relative evolutionary distance of *LSAMP* between the organisms shown. The line-length indicates the relative distance. Produced by ClustalW (Larkin et al., 2007).

The ability to use such a wide range of organisms is because all living organisms are descendants from a common ancestor and the genetic material and metabolic pathways of these have been conserved, to a greater or lesser degree, over evolutionary time. During the work described here chicken will be used as the model organism.

1.5.1 Mouse as an experimental model: Advantages and Disadvantages.

Using mouse as an experimental model imparts a number of advantages to the investigator. It is a well-established practice, with well documented techniques (Cold Spring Harbor Protocols, 2006, Lawson, 2000, Nagy, 2003) (Wassarman and Soriano, 2010, Hedrich, 2012) and a fully characterised genome (Mouse Genome Sequencing et al., 2002, Church et al., 2009).

Although it varies between species, birds and mammals have brains that are between six and ten times as big as the brains of reptiles of the same body size (Northcutt, 2002). Chicken-type birds and rodents have comparatively the smaller brains (Nieuwenhuys et al.,

1998). All human homologous brain areas and cell types are present (Bier and McGinnis, 2004). Reverse genetic engineering by gene knock-out (Capecchi et al., 2007) and site directed mutagenesis techniques are available (Thomas and Capecchi, 1987). Chimeric embryos construction is possible (Tam and Rossant, 2003) and a large mutant mouse collection is available (Wilkinson et al., 2010).

There are however some disadvantages to using mouse, mainly in connection with accessing the embryo. A very condensed protocol for generating transgenic (Cho et al., 2009) offspring would be as follows:

1. Immature donor females are super-ovulated by injection with PMSG (pregnant mare's serum gonadotropin) and HCG (human chorionic gonadotropin) and mated with stud males.
2. 12 hours post coitus the oviducts from the killed donor females are collected and the eggs harvested and placed in culture
3. Eggs microinjected with purified cloned DNA or plasmid.
4. Surviving eggs are implanted in pseudo-pregnant females (asexual mating with vasectomised males and then screen for plugging). These mice are anesthetised and following the surgical exposure of the infundibulum 15-20 eggs are injected into the ampulla of the oviduct. The oviduct and ovary are tucked back into the cavity and wound closed up. This is repeated on the other side of the mouse. The30-40 eggs implanted should give 5-10 pups depending on skill level.
5. Following birth at ~20 days, or earlier embryo extraction, the embryos can be processed depending on requirements e.g. further mating or staining for reporter vector.
6. If a germline is being developed the pups should be genotyped to identify transgenic mice. These mice are the F0 or founder mice.

7. Once the founders reach sexual maturity, they can be crossed with wild-type mice, preferably from the same strain as the F0 mice, to establish separate transgenic lines.
8. After verifying transmission from founder mice (F0) to offspring (F1), transgene expression within each line of the F1 or F2 generation has to be examined in most of the organs to prove that the line has proper temporal and spatial gene expression patterns and no leaky expression.

Another problem is that some early acting phenotypes are difficult to study because they are embryonic lethal (Xie et al., 2005), further embryonic manipulation, following reimplantation, is difficult as the embryo is inside the mother. Development and life-cycle of the animals is relatively slow (Bier and McGinnis, 2004).

1.5.2 Chickens offer Advantages and Disadvantages as a Model Organism in comparison with Mice.

The protocol for generating transgenic mice quite complicated, time-consuming and expensive. In contrast using chick is as the model organism is faster and easier because the embryo is accessible *in ovo*, the chick embryos do not have to be implanted and chicks mature quicker, for example 2 days development in chick is equivalent to about 9.5 days in mouse (Butler and Juurlink, 1987), allowing for a faster turnaround. From a cost basis, chickens are ideally suited as a model because fertilised eggs are cheap, costing approximately 33 pence each and are easily obtained from commercial hatcheries that produce large amounts of fertilized eggs for the pharmaceutical industry or agriculture. Maintenance costs are extremely low, there being no animals to feed or house by the experimenter and the equipment used for storing and incubating the eggs are widely used in poultry farming.

There are nonetheless disadvantages to using chick as against mice, for example the chick genome (International Chicken Genome Sequencing, 2004) hasn't been as well characterised as the genome of the mice typically used in laboratories. Initially generating transgenic chickens proved difficult (Dodgson, 2003, Siegel et al., 2006) however the protocol is now better established (van de Lavoir et al., 2006, Chapman et al., 2005, Mozdziak et al., 2003, Koo et al., 2006, McGrew et al., 2004)

However, because evolution in the genome doesn't follow a single linear pathway, each branch evolves separately at different mutation rates this leads to leading to some anomalies. For example although the chick genome is less similar to human than the rodent genome; the last common ancestor shared between the human line and the chicken line is 230 MYA whereas the last common ancestor with rodent is 75 MYA, it must be noted that the rodent genome has mutated more rapidly than the human genome (Li and Wu, 1987) and also that there are variations in the mutation rate across genomes both of which pose their own challenges (Hodgkinson and Eyre-Walker, 2011).

1.6 Analysis of gene expression by reporter vector.

Many diseases are the result of abnormalities in genes or chromosomes which can be caused by a mutation in a single gene and be Mendelian, or be more complex and be caused by mutations in several genes i.e. be polygenic. Lifestyle and environmental factors may also contribute to the development of the disease. Because of the complexity of the polygenic and multifactorial the identification of causes of Mendelian diseases is much more advanced. One of the many diseases caused by genetic mutation is Sickle cell disease (SCD), also known as sickle cell anaemia (SCA).

SCD is the result of a point mutation in the haemoglobin beta gene (HBB) that causes the amino acid valine to replace glutamic acid in the HbA β chains and is most prevalent in sub-

Saharan Africa, India, Saudi Arabia and Mediterranean countries (World Health Organisation., 2006). Mouse models have been generated to study SCD. One such is the homologous knock-in, which replaces the mouse globin genes with human genes (Wu et al., 2006).

Cystic fibrosis (CF) is an autosomal recessive genetic disease that is frequently fatal, and is more prevalent in Caucasians than other ethnic groups. It causes the body to produce thick mucus that clogs the lungs, leading to infection, and blocks the pancreas, preventing digestive enzymes from getting to the intestines. Since the identification of the gene responsible, the cystic fibrosis transmembrane conductance regulator gene (CTFR) (Rommens et al., 1989, Riordan et al., 1989, Kerem et al., 1989) many CF-causing mutations associated with CFTR. However, the leading cause of cystic fibrosis is produced by having the CFTRΔF508 mutation (Bobadilla et al., 2002) which is a deletion of the three nucleotides that code for phenylalanine (F) at position 508 (Kerem et al., 1989).

Huntington's chorea or disease is an autosomal dominant genetic neurodegenerative movement disorder is caused by multiple repeats of a CAG triplet at the 5' end the *Huntingtin* gene (Macdonald et al., 1993) resulting in a toxic gain-of-function. The CAG triplet codes for the amino acid glutamine. Typically there are up to 35 repeats of the triplet, but above that number of copies the phenotype becomes evident with higher number of repeats associated with an earlier onset age (Ravina et al., 2008, Andrew et al., 1993).

Analysing the effect of gene mutation and its associated expression has a long history and many methods have been used. One method is to start at the disease level and search for genetic differences between the affected population and the healthy population. This has become more automated now with the deciphering of the genetic codes for various organisms and several humans and the advent of gene-wide association studies (GWAS).

Another method is to observe the result to the phenotype when a gene is knocked out or indeed, knocked-in. Due to redundancy gene knock-out sometimes will have little or no effect (Perez-Perez et al., 2009). Conversely, due to the same gene being used many times in development the knock-out of certain genes may result in multiple phenotypic changes and so be lethal. This however can be overcome by using tissue-specific enhancers that drive site-specific recombinases in specific tissues to knock out the target gene in a very explicit manner (Hirano et al., 2011, Kos, 2004). A similar method uses tissue-specific enhancers to drive a toxin that destroys certain cells (Lee et al., 2000). There are also many other methods such as the Cre-Lox recombination system (Sauer and Henderson, 1988, Sauer, 1987). This system consists of the Cre recombinase enzyme that recombines a pair of short target sequences called the Lox sequences. The Cre protein cuts the DNA at both Lox sequences sites by the Cre protein and the strands are re-joined by DNA ligase. Inverted Lox sites will cause an inversion of the intervening DNA, while a direct repeat of Lox sequences will cause deletion of the intervening DNA. Translocation of the DNA to another chromosome can also be arranged as can triggering by chemical or heat-shock. A system with similar function is the FLP-FRT (Zhu and Sadowski, 1995, Turan and Bode, 2011, Schlake and Bode, 1994) but this uses the recombinase (FLP) on the flippase recognition target (FRT).

There are also the "Tetracycline-Controlled Transcriptional Activation Systems", methods of inducible gene expression where transcription is reversibly turned on in the presence (Tet-on) (Gossen et al., 1995) or absence (Tet-off) (Gossen and Bujard, 1992) of the antibiotic tetracycline, doxycycline or similar (Gossen and Bujard, 2002).

None of the above methods were used in the development of this thesis; instead a closed reporter vector method was used. In this method a putative enhancer sequence is inserted upstream of a minimal promoter and a chromogenic or fluorescent reporter gene. If the enhancer is activated by intercellular signals the promoter causes the expression of the

reporter gene and from this the temporal and spatial expression pattern of when and where an enhancer is active can be ascertained. This reporter system is usually introduced into the host organism in the form of a plasmid. Plasmids are circular DNA molecules that are separate from the DNA of the host cell, and typically remain so and therefore become more dilute at each cell-division, although some may undergo restriction enzyme digestion and become incorporated into the host's genome. Plasmids are very useful in the laboratory because they can be extracted from their host cell, manipulated by, for example, restriction enzymes that will open the plasmids and so allow the insertion of foreign DNA into the plasmid. Following ligation, where the circular plasmid is reformed containing the foreign DNA, they can then be inserted into a different cell and express the mRNA coded-for by the inserted DNA. At this stage they are known as vectors. They are known as expression vectors where they are constitutively active or reporter vectors where they only become active in response to a specific event or signal.

In the case of a eukaryotic vector a number of regulatory sequences are required in the plasmid for it to function as a vector. The most important is the transcription promoter sequence to allow assembly of the RNA polymerase and its associated transcription machinery. A polyadenylation sequence to insert the polyA tail on the mRNA is also needed. An enhancer sequence is similarly required to increase transcription starts above a low, basal level. For an expression vector, which will be constitutively on, the promoter/enhancer combination should be strong and the enhancer should respond to signals that are constantly expressed by the host cell. One such promoter is the β-actin promoter that in its native state is responsible for the expression of the gene ACTB. ACTB codes for a cytoskeletal actin that, being vital for the structure and integrity of eukaryotic cells is ubiquitously expressed. Another promoter that is constitutively active and strong in many infected eukaryotic cells is the cytomegalovirus (CMV) immediate early promoter (Boshart et

al., 1985) . This makes CMV useful as a promoter in plasmids but this can vary in expression and can give poor results in chick where expression is largely confined to the pancreas, although it does function at some level in most tissues (McGrew et al., 2004).

Where a reporter vector is required, the needs for the promoter-enhancer combination differ somewhat, depending on the circumstances. When testing whether an enhancer is being activated or not it is vital that promoter only becomes active in conjunction with this activated enhancer. The promoter must on no account be 'leaky', i.e. allow transcription of the reporter gene to start without the signal from the activated enhancer. Some reporter vectors use a minimal promoter to provide this functionality.

A minimal promoter consists of the few nucleotides in those sequences that are absolutely required for binding of the transcription factors to cause transcription of the gene. The human β-globin minimal promoter region contains CAC, CAAT and TATA box regions it also binds the NF-E1 (also known as GATA-1) nuclear factor and ubiquitous factors (deBoer et al., 1988).

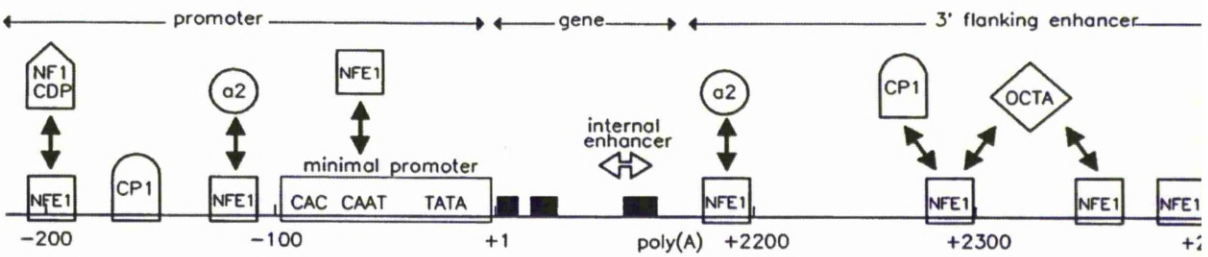


Figure 3 Schematic representation of the factors binding to the human ,β-globin promoter (deBoer et al., 1988) and 3' enhancer (Wall et al., 1988). The promoter and enhancer are on the same scale, the gene and immediate 3' flanking region have been compressed to a smaller scale. The symbols indicate the various protein factors. CDP, NFl, NF-E1, CPI, a2 and an octamer binding protein, b3/c2 Wall 1988 (Wall et al., 1988). Note that the factors binding to the minimal promoter (TATA, CAAT and CAC boxes) have not been indicated. Solid double arrows indicate binding of different factors to the same or overlapping sites. (deBoer et al., 1988).

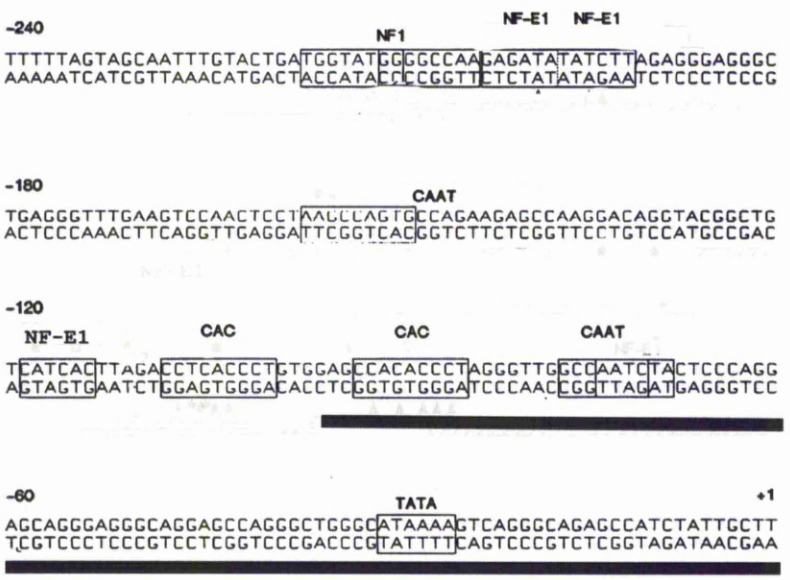


Figure 4 The human βGlobin gene layout. The minimal promoter k black bar and runs in the region -100 to +1. The boxes indicate consensus sequences TATA, CAAT, CAC, NF-El and NFl. Adapted from Fig 2 (deBoer et al., 1988).

1.7 Transfection of selected Tissue with Reporter Vector.

In order to function the reporter vector has to be transfected into the cells of the organism under investigation. There are many different ways of transfecting cells ranging from chemical (e.g. Lipofection), through Biolistic (particle-based) to viral-based.

Of three non-viral methods investigated to transiently transfect chick eggs *in-ovo* (Lipofection, Biolistics and Electroporation), electroporation was found to give the highest survival rate of 53.6% (30/56) and highest transfection rate 50.0% (15/30) (Muramatsu et al., 1997) . Electroporation or electro-transfection, is popular because it works on most types of cells, is cheap, cloning can be directed to a particular location such as the neural tube , it can be efficient with many cells taking up the DNA (Miller and Nickoloff, 1995), had been successful in neural tube (Timmer et al., 2001) and of particular interest, can be used *in-ovo* into chick (Muramatsu et al., 1997, Momose et al., 1999). All of these methods are being constantly improved, but at the time electro-transfection was deemed to be the most suitable and was used as the transfection method during the preparation of this thesis.

Electro-transfection consists of two main stages, the electroporation or electropermeabilization stage where the cell's membrane is made porous and the electrophoresis stage where the DNA is moved towards the membrane (Figure 5).

Electroporation is achieved by placing an electric field sufficient to cause a potential difference of a critical value across the cell. When this critical value is reached, molecules, which under normal conditions would not be able to penetrate the cell's membrane, are able to do so. It is not fully understood what events occur within the membrane to facilitate this. One theory is that the electrical field increases surface tension thus altering the balance of the two opposing forces of surface tension and line tension. As surface tension favours pore formation the opposing force of line tension is overcome thus pore formation is

favoured (Weaver, 1995). An alternative theory suggests that the electric field compresses the membrane bilayer, while a further theory is that polar head group of the phospholipids is reorganization resulting in a weakening of the hydration layer (Stulen, 1981, Lopez et al., 1988). Whichever of these theories eventually proves correct the electroporation is asymmetrical with greatest effect being on the sides facing the electrodes and being most pronounced on the side facing the anode (the positive electrode) (Hibino et al., 1993).

The critical value to allow poration, at a particular point on the cell, is typically between 200mV-600mV (Teissie and Rols, 1993, Gabriel and Teissie, 1997) and is independent of the cell type. The electric field required to achieve this critical value is inversely dependant on cells size, bacteria require up to 200 kV/cm whilst large cells only require 200V/cm (Teissie and Rols, 1993). Electrically coupled cells such as those in developing chick neural tube (Sheridan, 1968) behave electrically as if they were a single large cell thus reducing the voltage required for poration.

A further complication is that once the critical field strength has been reached, it is the duration and number of electric pulses that determines the extent of poration, increasing the field strength only increases the area where the poration occurs (Gabriel and Teissie, 1997). Too high a field strength will results in irreversible poration or Joule heating, both resulting in cell death. A square wave pulse is now more favoured than that with an exponential decay as it reduces the amount of time the electric field is applied and thus reduces the Joule heating and risk of damage i.e. burning.

Electrotransfection's disadvantages stem mainly from this risk of cell damage and the non-specific transport, of for example, calcium and sodium into the cell and, for example, potassium, out of the cell when the membrane has been permeabilized.

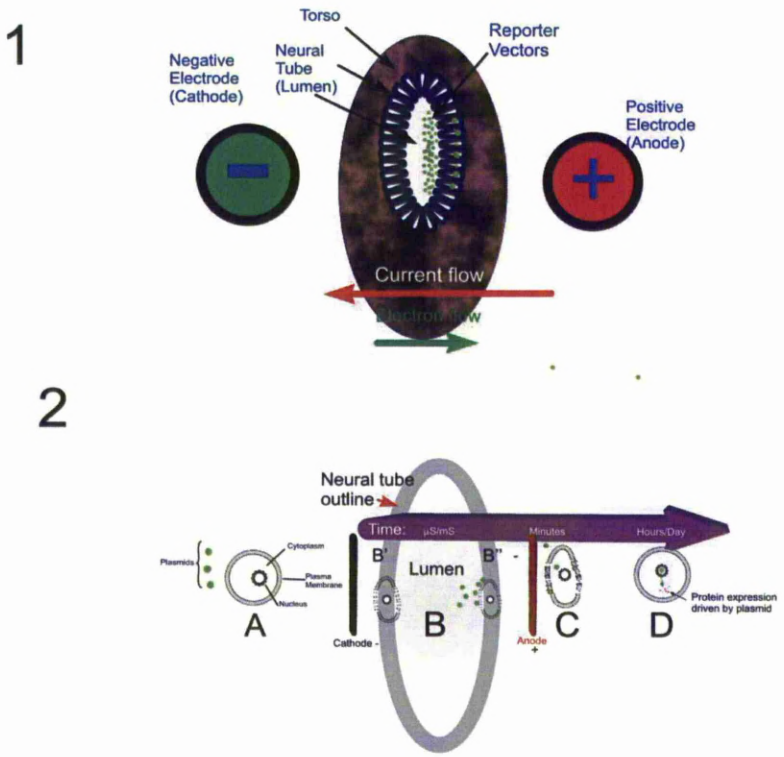


Figure 5 Electro-transfection process.

(1) Cross section through embryonic torso showing cells of the neural tube and its lumen. At this stage plasmid has been injected into neural tube and current pulses applied across the embryo from the two electrodes. Plasmid (green dots) are shown attracted by the anode (+) by electrophoresis and some of the cells are shown as transfected with plasmid. This coincides with image 2C. The direction of the current flow from anode to cathode is indicated with a red arrow whilst the electron flow from cathode to anode is shown with a green arrow. (2)Timeline of electroporation: (A) Prior to application of electric field the cell is shown as circular with plasma membrane, cytoplasm and nucleus marked. Plasmids in the extra-cellular space are shown as green circles. The plasma membrane is intact and the cell controls most transport in and out of the cell. (B)The neural tube is outlined in grey with two representative cells B' and B'' shown. During application of electrical field the cells (B' and B'') are shown as deformed and pores have opened within the cell's plasma membrane on the side nearest the anode (shown as red bar) which is now hyperpolarised, and to a lesser extent on the side facing the cathode (shown as black bar), which is depolarised. The plasmids have undergone electrophoretic migration towards the anode and so away from cell B', and some are in intimate contact with both the intact and permeated parts of the membrane of the cell near the anode (B''). (C) Electric field is removed and plasmids at the permeated part of the membrane are translocated into the cytoplasm. The plasma membrane has started to regain its integrity but the cell is still at risk of non-specific migration of molecules into and out of the cell. (D) Several hours later the membrane has further regained its integrity and the plasmids have migrated intracellularly and are now present in the nucleus, probably as a result of mitotic activity and constitutively active plasmid should express their protein and be detectable within a day.(2) adapted from*(Escoffre et al., 2009)*

1.8 Long-term Anxiety is debilitating and costly.

Anxiety and fear are familiar traits in humans and animals and in the correct context increase survival (Olsson and Phelps, 2007, Marks and Nesse, 1994) . However, many people in the modern, relatively safe environment, exhibit levels of these emotions that are maladaptive and interfere with normal functioning.

These anxiety disorders result in behavioural effects that may by diagnosed as one of the well characterised conditions such as panic attack, post traumatic panic disorder, generalised anxiety disorder or social phobia (Kessler et al., 2005).

Indeed, anxiety disorders are now considered a collection of a number of associated disorders incorporating fearfulness, emotional instability and stress reactivity that include general anxiety disorder (GAD), panic disorder, agoraphobia, social phobia, specific phobia, obsessive-compulsive disorder (OCD), acute stress disorder and post-traumatic stress disorder (PTSD) (American Psychiatric Association, 2000).

The personal cost of these anxiety-related disorders is enormous and so too is the cost to the economy due to loss of productivity. In England, the total number of people with anxiety disorders was estimated to be 2.28 million in 2007 however only 49% of these are in contact with services and only 46% of these receive treatment. The cost of treatment was £1.2 billion and when loss of earnings, for those under 65, of £7.7 billion is added to this the total cost was £8.9 billion (King's Fund et al., 2008). If 95% of the current sufferers were treated the treatment cost would rise to £2.1 billion but the loss of earnings should reduce (King's Fund et al., 2008). Interestingly, presenteeism from mental illness costs the UK economy almost double the cost of absence due to mental illness (Sainsbury Centre for Mental Health, 2011). The number of those suffering from anxiety disorder is predicted to increase

due to demographic changes in age and ethnicity. This will cause the cost of treatment to rise to an estimated £2.04 billion and loss of earnings to £12.15 billion in 2026.

Large though these figures are they may be underestimated as they are based on anxiety disorders in the community being 5.4% but other surveys showed higher rates, the United States National Comorbidity Survey Replication returned an 18.1% rate (Reviewed in (Baumeister and Harter, 2007)). Another factor is the age-related rate which increases with age, reaching a maximum in the 45-54 age group and then reducing after that. This reduction may only be an artifact because other conditions may hide anxiety disorders (Alwahhabi, 2003) and a Dutch study found the rate for older people (55-85 years old) to be 10.2% (Beekman et al., 1998).

A number of treatment regimens are available, either alone or in combination, including Cognitive Behavioural Therapy, Psychotherapy/Counseling and drug therapy. Drug therapy can include β-blockers, benzodiazepines, anxiolytics and antidepressants. All of these have greater or lesser side effects ranging from the high potential of addiction to benzodiazepine, the less likely heart failure associated with β-blockers to the annoying dry-mouth that is a side effect of most of these medications (Papadakis et al., 2013, Cecil et al., 2004, Longo and Harrison, 2012).

Some of these medications are only marginally better or no better than placebos. Two Japanese double-blind trials found placebos as effective as Buspirone (Okada et al., 1993) although Buspirone is still prescribed for anxiety in the UK (MIMS, 2012) .

In England, of the 54% of patients treated, varying rates of success from the different treatment regimens have been reported (Table 1) with combined drug and counselling treatment being the most effective, however the report did not specify if there was any recurrence of the condition.

Treatment	Patients/%	Health Improves/%
Medication only	35	54
Psychological Therapy only	09	46
Combined Treatment	11	72
No treatment	46	26

Table 1 Treatment and success obtained among people with anxiety disorders in contact with mental Health services in England. Data from (King's Fund et al., 2008).

Diagnosing anxiety and many other mental illnesses is highly subjective, relying on the patient responding to a questionnaire often based on the latest Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000) and the interpretation and diagnostic skills of the clinician. Given that the patient may be nervous due to the clinical setting and may be presenting with comorbidity, either physical (Orhan and Uluşahin, 2001) or mental (Freeman et al., 2002), isolating anxiety as a diagnosis may not be straightforward . A genetic or biomarker test would help remove the subjectiveness inherent in the current system and so inform the diagnosis and subsequent treatment, thus reducing the risk of an inappropriate treatment regimen with its associated iatrogenic risks.

One potential genetic test would be of the expression rate of certain genes that are implicated in mood and anxiety.

1.9 We are all different due to Genetic and Genomic variations

The genome of each advanced organism differs due to germ-line and somatic mutations and/or polymorphisms. There is no biochemical difference between a mutation and a polymorphism; they differ only in frequency of occurrence. A mutation is defined as a

polymorphism when it affects at least 1% of the population at which stage no single allele is deemed to be the standard sequence, but instead there are a number of equally acceptable variants (Harris, 1969).

Depending on what the variation in DNA is, and where it occurs, defines if there will be any phenotypic effect. If the variation occurs in an exonic region it is possible that there will be a variation in the protein that is being coded for. When the mutation occurs in a non-coding DNA sequence, there will probably not be any effect on any protein sequence; however there may be more subtle effects such as rate of protein expression. It is possible that the sequence change may produce a synonymous codon and produce no effect whatsoever.

These mutations have been classified into a number of categories that includes; single nucleotide polymorphism (SNPs); insertion-deletions(INDEL); variable number of tandem repeats (VNTR, minisatellite (Jeffreys et al., 1985, Fondon and Garner, 2004)); Single Tandem Repeats (STR ,microsatellite (Turnpenny and Ellard, 2005, Turnpenny et al., 2012)); copy number variants (CNV (Sebat et al., 2004, Korbel et al., 2007)) and repetitive elements e.g. Alu, LINES, SINES (Darling et al., 1982, Schmid and Deininger, 1975, Singer, 1982)

VNTRs are a short sequence of DNA (20-50 nucleotides) that is repeated a variable number of times (Jeffreys et al., 1985). These are not uniformly distributed through the genome but tend to be concentrated at telomeres and in hypervariable regions. STRs are similar to VNTRs but are very short sequence of DNA (2-6 nucleotides) that are also repeated a variable number of times (Nakamura, 2009).

CNVs describe the gains and losses of large blocks of DNA sequence (10K - 5000K) and are implicated in vulnerability and resistance to disease whereas a SNP defines when a single nucleotide is exchanged for another (A,T,C or G) resulting in a DNA sequence that differs between members of a biological species or paired chromosomes in an individual.

Various human-based association studies have shown that there is a link between polymorphisms in genes and their control regions that affect, for example, the serotonergic system (Inada et al., 2003), the dopaminergic system (Maron et al., 2005), G-protein signalling regulation (Koenen et al., 2009) and modulation of monoamine metabolism (Domschke et al., 2007).

If a system were available to rapidly screen these control regions a clinical tool could be developed. This thesis describes the initial development of this methodology and covers the selection of a gene to investigate, identification of putative control regions of that gene and the use of chick as a model organism.

1.10 Limbic System Associated Membrane Protein gene (*LSAMP*) has been implicated in Anxiety and Mood Disorders.

Association studies of genetic variations in *LSAMP* have shown SNPs to be implicated in panic disorder (Koido et al., 2006, Maron et al., 2006). A report compared DNA from 288 male suicide victims and 327 healthy male volunteers conveyed a link to suicide (Must et al., 2008) appears to be tenuous at best. SNP (rs7632246 A/C) of the *LSAMP* gene is related to panic disorder (Koido et al., 2006, Maron et al., 2006). *LSAMP* has also been identified in two rodent-based studies as an important regulator of behaviour. In the first of the rodent-based studies, *wt* Wistar rats were classified as anxious, intermediate or non-anxious, depending on their exploratory activity in an elevated plus maze (Nelovkov et al., 2003). Although Wistar rats are an outbred strain, which should result in genetic variation the regimen under which they are allowed to breed can result in differing genetic variation between colonies. Following comparison of the gene expression between the anxious and non-anxious rats 96 differences were found, but Nelovkov highlighted increased *LSAMP* expression as being the most intriguing. Further work (Catania et al., 2008) demonstrated

that the genetic deletion of *LSAMP* in mice causes hyperactivity in novel environments. In generating the knock-out mice, one heterozygous *LSAMP* founder was obtained and back-crossed into the C57BL/6J strain for more than 10 generations for all their experiments.

The knockout mice showed a higher exploratory drive than their *wt* counterparts, which is taken to be compatible with lower anxiety levels. These findings indicate that different expression levels of *LSAMP* may be involved in certain mental dysfunctions. Changes in *LSAMP* expression are therefore relevant to anxiety. *LSAMP* has also been implicated in mental health issues such as schizophrenia (Behan et al., 2009), male completed suicide (Must et al., 2008), panic disorder (Maron et al., 2006), and mood and anxiety disorders (Koido et al., 2006). Problems resulting from *LSAMP*'s expression are not confined solely to affective disorders. Mutations in *LSAMP* that reduce its expression are implicated in left main artery disease. It is purported that the lower levels of LAMP in the arterial wall increase the production of smooth muscle cells possibly resulting in worsened atherogenesis (Wang et al., 2008).

Given the evidence that *LSAMP* expression is implicated in anxiety and mood and the potential economic savings that would accrue if an effective diagnostic tool could be developed, *LSAMP* was selected as the first gene to be investigated.

1.11 Limbic System Associated Membrane Protein (LAMP) is a member of the Immunoglobulin Superfamily (IgSF).

LAMP is a member of the IgLON family of the Immunoglobulin Superfamily (IgSF) (Pimenta et al., 1995). Immunoglobulins are probably best known for their role in antibodies of the immune system where they can discriminate between auto-generated proteins, foreign,

unwanted proteins and other antigens. Immunoglobulins are also involved in cellular recognition and adhesion processes. This ability to recognise and bind specific complementary elements is provided by a domain known as an immunoglobulin domain or fold (Bork et al., 1994, Brummendorf and Rathjen, 1995). The immunoglobulin fold consist of two anti-parallel arrays of β-strands that form two β-pleated sheets. The two sheets are joined together by disulphide bonds (Figure 6).

By definition, each member of the hundred-strong immunoglobulin superfamily (IgSF) contains at least one Ig-like domain. Members of this superfamily are commonly associated with roles in the immune system such as cell surface antigen receptors, antigen presentation to lymphocytes and cell adhesion molecules (Table 1). IgCAMs form the largest family of calcium-independent cell adhesion molecules in vertebrates and some act as signal receptors that transmit information from external sources to the signalling systems within the cell (Brummendorf and Rathjen, 1995).

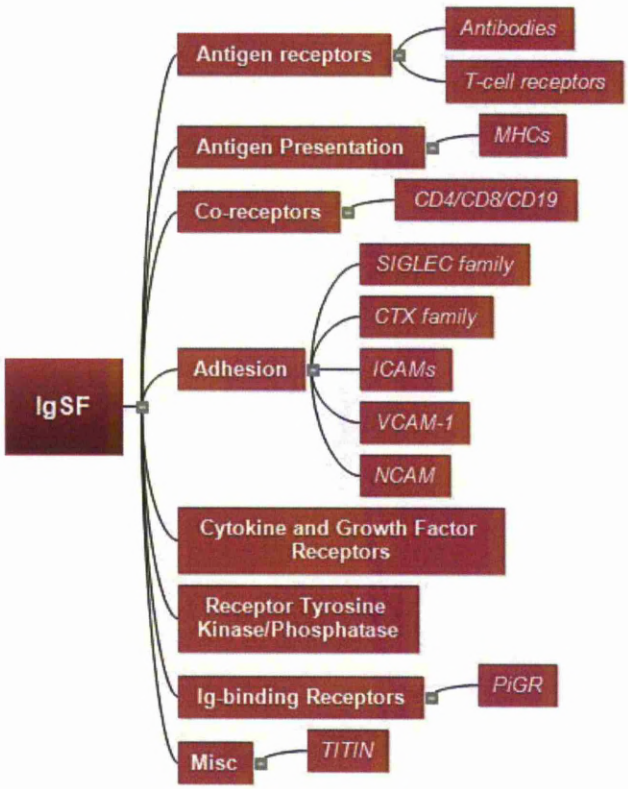


Table 2 IgSF functions. Example of the function of the IgSF and some of the members associated with those function. Data from (Walmod et al., 2007) (Barclay, 1999) (Rougon and Hobert, 2003) (Buck, 1992)

1.12 An IgCAM must contain at least one Ig-like domain.

The Ig-like domains of the IgCAMs range from 70 to 110 a.a. and are typically formed by two anti-parallel β sheets packed face-to-face. The sheets themselves consist of between 3 and 5 anti-parallel β-strands designated A through G. Each strand is 5 to 10 amino acids long joined by loops (Williams and Barclay, 1988) of variable lengths (Brummendorf and Rathjen, 1995). The cysteine residues contained by most Ig-like domains that form the disulphide bridge occur on strands B and F, about 55 to 75 amino acids apart (Williams and Barclay, 1988). The Ig-like sequences had been categorised initially into C1, C2 and V sets (Williams

and Barclay, 1988) and later an I set, that differs in the count and length of the β-sheets and length and structure of the loops was described. The V and C1 are Variable and Constant as with immunoglobulins, C2 combines characteristics of both, having the smaller loop of the C1-type but some amino acid sequence towards the C-terminal similar to V. The IgSF domains of lower organisms are found in the I set (Bateman et al., 1996, Harpaz and Chothia, 1994).

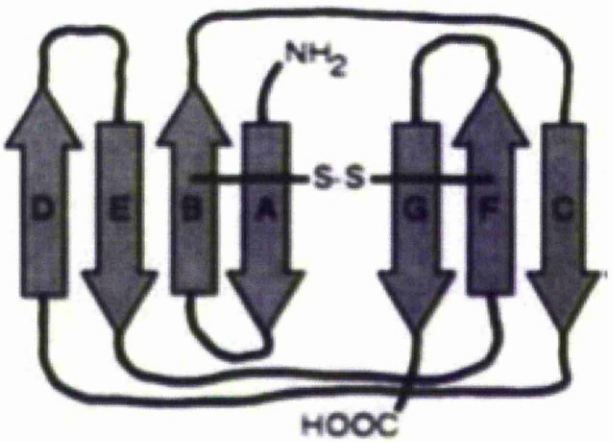


Figure 6 Diagram of strand topologies in IgF glycoprotein. One ,β-sheet of Ig type C2-like domains contains strands A. B. E. and D, while the opposing sheet contains strands G. F, and C. (Adapted from (Williams and Barclay, 1988)).

As previously mentioned IgCAMs must have at least one Ig-like domain; P_0 the smallest, possess only a single Ig-like domain (Figure 7) , but as well as the Ig-like domains IgCAMs can also include other protein domains (Figure 7) such as Fibronectin III (Brummendorf and Rathjen, 1995). IgCAMs can exhibit heterophilic binding, i.e. bind to non-identical adhesion molecules as well as homophilic binding, i.e. bind to identical molecules as themselves (Maness and Schachner, 2007). They may have a cytoplasmic transmembrane region, be

GPI-anchored and so possess no transmembrane region (Maness and Schachner, 2007) or lack a GPI-anchor and thus be soluble (Lodge et al., 2001b). As well as having a binding function some IgCAMs can communicate intracellularly. For example some forms of N-CAM associate with Src family cytoplasmic tyrosine kinases which relay signals by phosphorylating intracellular proteins on tyrosines (Beggs et al., 1997, Paratcha et al., 2003, Maness et al., 1996). Other Ig family members are transmembrane tyrosine phosphatases that help guide growing axons to their targets by dephosphorylating intracellular proteins (Angers-Loustau et al., 1999, Sallee et al., 2006, Burridge et al., 2006, Bouyain and Watkins, 2010, Revest et al., 1999) . IgCAMs L1 and NCAM have been shown to act as signalling co-receptors in neuronal migration and signalling outgrowth (Schmid and Maness, 2008).

The extracellular portion of some IgCAMs can be cleaved off by 'a member of the disintegrin and metalloprotease' family (ADAM) and 'matrix metalloprotease' (MMP) and shed resulting in soluble ligands which can interact with receptors. An example of this is given by the Junctional Adhesion Molecule C (JAMC) which appears to promote angenesis (Rabquer et al., 2010). Upon cleavage the intracellular portion can be released into the cytosol and can influence cellular functions. When the extracellular portion of L1 is cleaved the intracellular fragment is released from the membrane by γ-secretase and then enters the nucleus and influences the transcription of certain genes, for example β3-integrin and cathepsin-B (Riedle et al., 2009).

Two members of the IgSF, Sdk-1 and Sdk-2, are expressed by separate sets of neurons that subdivide the dendrites into narrow sublaminae in the chicken retina. The cell adhesion molecules are located on synapses and appear to use homophilic binding to direct formation of explicit contacts (Yamagata et al., 2002).

When forming dimers or multimers in *cis* or in *trans* it is understood that many IgCAMs remain fully extended, however some adopt what is referred to as a 'horseshoe' conformation i.e. the Ig-like domains at the N terminus bend over. One example of this is neurofascin (Liu et al., 2011) where the outermost Ig-like domains bend over. Another IgSF member, consisting of two Ig-like domains and a transmembrane region, that adopts a horseshoe formation is the coxsackievirus-adenovirus receptor (CAR). This it appears to do when it homophilically interacts in *cis* (Patzke et al., 2010), when binding in *trans* it appears to assume a more vertical stance.

The role of the IgCAM, 'Close Homologue of L1' (CHL1) has been found to have the additional and unexpected role of being involved in the regulation of the uncoating clathrin coated synaptic vesicles (CCSVs). This is achieved by CHL1 associating with the constitutively active heat-shock protein Hsc70 and forming a complex with it in the plasma membrane (Leshchyns'ka et al., 2006). As CHL1 gene mutations have been linked to schizophrenia (Sakurai et al., 2002, Montag-Sallaz et al., 2003) this opens an interesting new path for investigation into IgCAMs and associated disorders.

1.13 LAMP is GPI-anchored to the cell membrane.

The most common membrane linkage of IgCAMs is a type I with a 25 amino acid α-helix hydrophobic region that spans the membrane (Beckerle, 2001) , however glycosylphosphatidylinositol (GPI) anchoring is not uncommon and is found in NCAM-120 (Leshchyns'ka et al., 2003) contactin,axonin-1 and BIG-1 and BIG-2 (Katidou et al., 2008) as well as the IgLON sub-family (Funatsu et al., 1999). LAMP is therefore GPI anchored.

1.14 IgCAMs differ in form and function.

IgCAMs consist of individual members and families ranging in decreasing complexity ranging from Sidekick which contains many extracellular domains (thirteen F3-homology modules

and six Ig-like molecules) to P_0 which contains only 1 extracellular domain (Figure 7). Necl proteins are an alterant name for SynCAM proteins (Biederer, 2006)

The different IgSF isoforms may be generated by alternate splicing of RNA transcript from a single gene, resulting in a large number of proteins with differing capabilities. Other modifications can also affect the function of the protein, for example the post-translational addition of polysialic acid of NCAM increases the range of inter-membrane repulsion negating homophilic NCAM and NCAM-mediated cadherin binding (Johnson et al., 2005).

IgCAMs with their wide variety of structures, expression patterns and interactions are ideal molecules to provide much of the extensive range of cues, both cell-cell and cell-ECM, required for a complex assembly to be formed from the building blocks of individual cells. IgCAMs have a role throughout the whole life of a multi-celled organism. During development the exquisite control required for neuronal induction, migration, correct target location and plasticity is provided by IgCAMs (Walsh and Doherty, 1997).

Due to their widespread expression some IgCAMs probably have a generalised role (Brummendorf and Rathjen, 1995) whilst others exhibit a specific expression pattern which indicates a tissue-specific role (Schwarz et al., 2009). The specific tissue and development stage regulates the type of variants generated by post-translation modification and alternative splicing (Walsh and Doherty, 1997, Hassel et al., 1997, Volkmer et al., 2012, Shapiro et al., 2007).

IgCAMs interact with other IgCAMs and proteins from other families indicating that they take part in a widespread network of control and interaction. Cadherins and IgCAMs are often found on the same cells and although cadherins exhibit the stronger binding that is required to segregate cells into tissues and maintain the tissue integrity, N-Cam and other IgSF may be involved in fine-tuning the adhesion in development. For example cadherin-

mediated aggregation (Dahl et al., 1996) and NCAM-mediated segregation (Esni et al., 1999) allow for the formation of the Islets of Langerhans. It has been shown in rodent that inhibition of cadherin function preventing aggregation results in the fatal result of no islet formation, whereas the inhibition of N-CAM allows for the formation of islets, albeit disorganised. It is hypothesised that non-junctional CAMs initiate cell-cell adhesions which are then orientated and stabilised by the organized intracellular junctions (e.g. desmosomes) (Alberts, 2002, Vestweber, 2000).

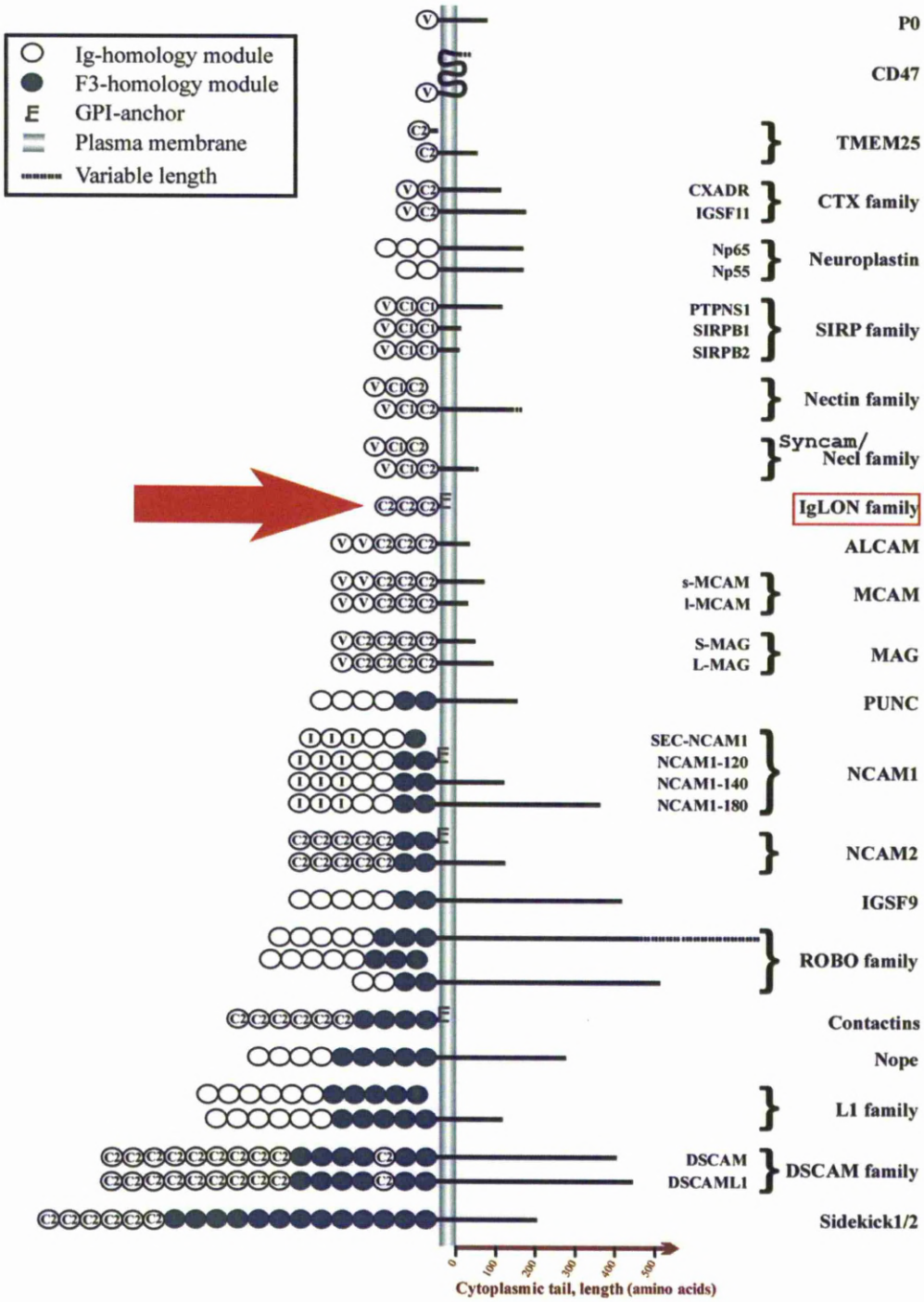


Figure 7 Ig-like CAMs in ascending order of complexity starting with those containing one Ig-like domain (P_0) and continuing on to those containing multiple Ig-like and Fibronectin3-like domains. The IgLON family is indicated by red arrow and box. Figure 2-1 (Walmod et al., 2007)

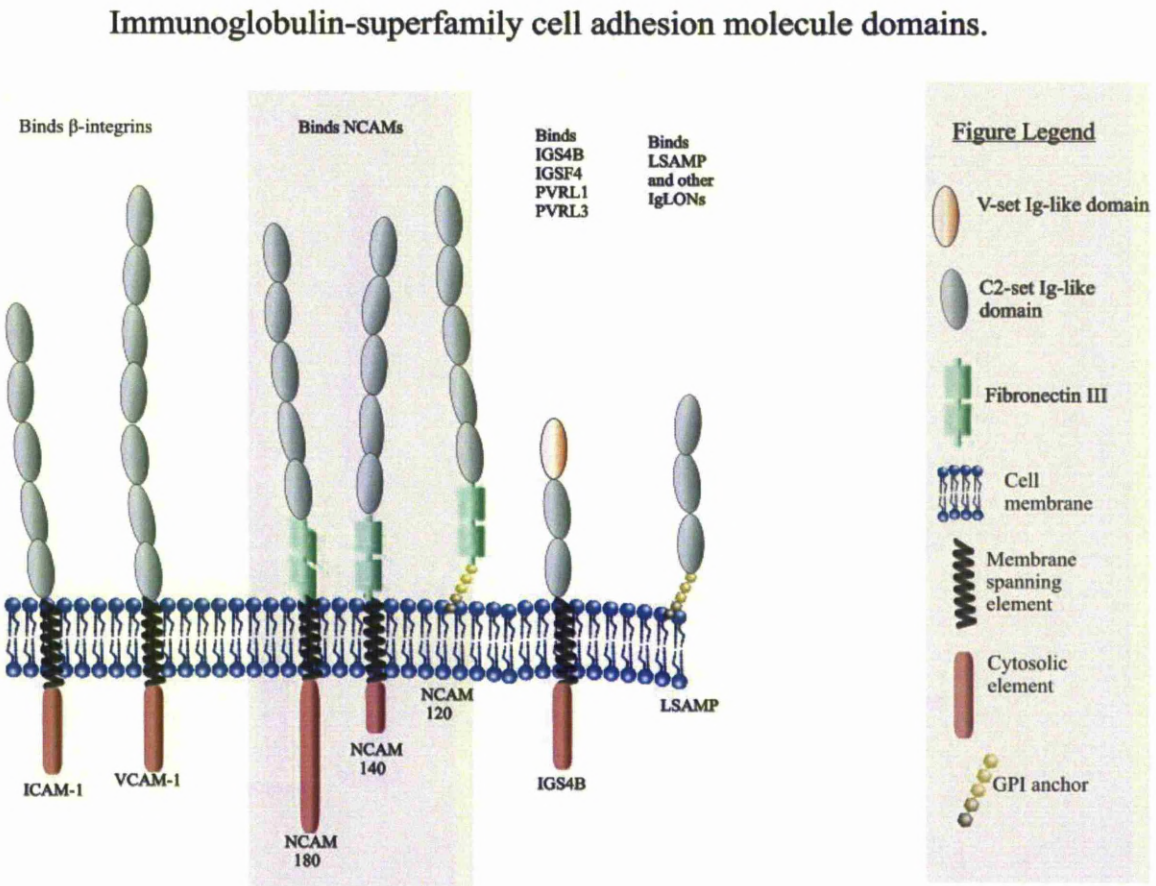


Figure 8 Examples of Immunoglobulin-superfamily cell adhesion molecule and their domains. Known binding interactions are also shown. Adapted from (Walmod et al., 2007)

1.15 IgLONs are a subfamily of the IgSF Superfamily

1.15.1 Introduction to IgLONs

IgLONs, members of the IgSF, are expressed in the nervous system during development and maturity (Akeel et al., 2011). During development they are probably involved in synaptogenesis, axonal outgrowth and differentiation of neuronal precursor in the brain and retina.

The IgLON subfamily consists of four main members (Table 2), *LSAMP* (Nakajima et al., 1997, Hancox et al., 1997, Wilson et al., 1996, Zhukareva and Levitt, 1995, Brummendorf et al., 1997), OBCAM (Schofield et al., 1989),Neurotrimin (Ntm/CEPU-1) (Struyk et al., 1995) and the last member of the family, Kilon (KILON/Ntr) (Funatsu et al., 1999, Marg et al., 1999, Schafer et al., 2005). The 'LON' in the name comes from the initial letters of the three original family members (Pimenta et al., 1995).

Originally it was purported that IgLONs functioned as axonal guidance molecules (Hachisuka et al., 2000, Pimenta et al., 1995), however LAMP was also described as facilitating axon target recognition during the formation of thalamocortical projections; LAMP repulsed non-limbic thalamic and cortical axons whereas, in a stripe assay, the limbic system axons preferred the LAMP substrate (Mann et al., 1998). More recently it was suggested that their most important role may not be that of axon guidance but that of cell–cell adhesion and recognition (McNamee et al., 2002).

1.15.2 IgLON structure

The IgLON structure typically consists of three Ig-like domains GPI-anchored to the cell's membrane (Figure 9) and located to lipid raft domains. However, CEPU-se, an isoform of CEPU1 (Lodge et al., 2001a), lacks the GPI anchoring domain and is therefore soluble. In chick, there is also a short version of KILON/Ntr that lacks the third Ig-like domain (Marg et al., 1999) . LAMP and Ntm/CEPU-1 both have β-isoforms containing an insertion between the third Ig-like domain and the GPI anchor (Pimenta and Levitt, 2004, Lodge et al., 2001a).

IgLONs exhibit a high degree of inter-sequence similarity (Table 3) with, for example in human, OPCML having 76% similarity with NTM (Goujon et al., 2010). There is also a high degree of inter-species similarity (Table 4) with, for example, an 85.% similarity between human and chicken *LSAMP* DNA, while the LAMP-protein's amino acid sequence shares 91.1% similarity. When human and chick's NTM is compared the DNA sequence shares 78.1% similarity and the resulting proteins, Neurotrimin and CEPU-1, are 80.4% similar (NCBI-Homologene, 2011). These data compare favourably with the earlier findings that the CEPU-1 protein had high similarity with the other IgLONs i.e. 76% with Ntm, 68% with OBCAM and 54% with LAMP (Spaltmann and Brummendorf, 1996).

The protein sizes of all the IgLONs in human are very similar with LAMP being 37393 Daltons (Da), OBCAM 38008 Da, Neurotrimin 37971 Da and Kilon 38719 Da (Magrane and Consortium, 2011) but as these proteins are highly glycosylated their apparent weight increases considerably. This difference was demonstrated in GLUT1 where its apparent molecular weight reduced to 38,000 Da from 55,000 Da with the loss of glycosylation (Asano et al., 1991) . A similar effect of glycosylation on the IgLONs would result in the apparent

55,000 Da found from the immuno-blots described in earlier literature (Struyk et al., 1995, Spaltmann and Brummendorf, 1996, Brummendorf et al., 1997).

Protein			Gene					
			Human			Chick	Rat/mouse	Abbreviation used in text of thesis †
Mammalian	**Chicken homologues**	**Abbreviation used in text of thesis †**	**Name**	**Symbol**	**Synonyms**	**synonym**	**symbol**	
LSAMP	*LSAMP*	*LSAMP*	limbic system-associated membrane protein	*LSAMP*	IgLON family member 3, IGLON3, *LSAMP*	chLAMP	*Lsamp*	*LSAMP*
OBCAM	OBCAM	OBCAM	opioid binding protein/ cell adhesion molecule-like	OPCML	IgLON family member 1, IGLON1, OBCAM, OPCM	OBCAM	Opcml	OPCML
Neurotrimin	CEPU-1 (Spaltmann and Brummendorf, 1996)	Ntm/CEPU-1	neurotrimin	NTM	HNT, IgLON family member 2, IGLON2, neurotrimin, NTRI	CEPU, CEPU-SE, HNT	Ntm	NTM
Kilon	Neurotractin (Marg et al., 1999)	KILON/Ntr	neuronal growth regulator 1	NEGR1	a kindred of IgLON, IgLON family member 4 IGLON4, KILON, MGC46680, neurotractin Ntra	NTRA-L, NTRA-S	Negr1	NEGR1

Table 3 Symbols and names of the IgLON family of CAMs showing their name in mammals, chicken, abbreviations used in thesis, their gene name, symbols and synonyms as approved for human by HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (HGNC) seal 2011 (Seal et al., 2011) ; chick synonym, Chicken Gene Nomenclature Consortium (CGNC) (Burt et al., 2009) ; rat symbols from Rat Genome Database (RGD, 2011, Shimoyama et al., 2011) and mouse data from Mouse Genome Database (Eppig et al., 2012)

† Symbols specific to an organism may appear in tables and material imported.

	OPCML	*LSAMP*	NTM	NEGR1
OPCML	--	54	76	46
LSAMP	--	--	55	55
NTM	--	--	--	45
NEGR1	--	--	--	--

Table 4 ClustalW comparison of percentage similarities between members of the IgLON family in human. OPCML/NTM exhibit the greatest similarity at 76% whilst NTM/NEGR1 share the least at 45%. (ClustalW using the following accession numbers - *LSAMP*, U41901;OPCML,NM_002545; NTM, NM_016522; NEGR1, NM_173808 (Goujon et al., 2010).

Species	symbol	%id protein	%id DNA
H.sapiens	***LSAMP***		
P.troglodytes	*LSAMP*	99.7	99.7
B.taurus	*LSAMP*	98.3	94.3
M.musculus	*Lsamp*	99.0	94.2
R.norvegicus	*Lsamp*	98.8	94.2
G.gallus	*LSAMP*	91.1	85.1
H.sapiens	**OPCML**		
P.troglodytes	OPCML	100.0	99.8
B.taurus	OPCML	97.7	90.6
M.musculus	Opcml	95.5	89.5
R.norvegicus	Opcml	98.3	92.7
G.gallus	OPCML	85.2	79.5
H.sapiens	**NTM**		
P.troglodytes	NTM	99.3	99.2
B.taurus	NTM	98.5	91.1
M.musculus	Ntm	98.1	90.0
R.norvegicus	Ntm	97.7	91.0
G.gallus	NTM	80.4	78.1
H.sapiens	**NEGR1**		
P.troglodytes	NEGR1	98.6	98.7
B.taurus	NEGR1	97.7	95.9
M.musculus	Negr1	95.7	88.8
R.norvegicus	Negr1	96.0	89.9
G.gallus	NEGR1	82.6	80.0

Table 5 IgLON inter-species protein and DNA sequence homology. All have been compared using Human as the base sequence. Data (NCBI-Homologene, 2011).

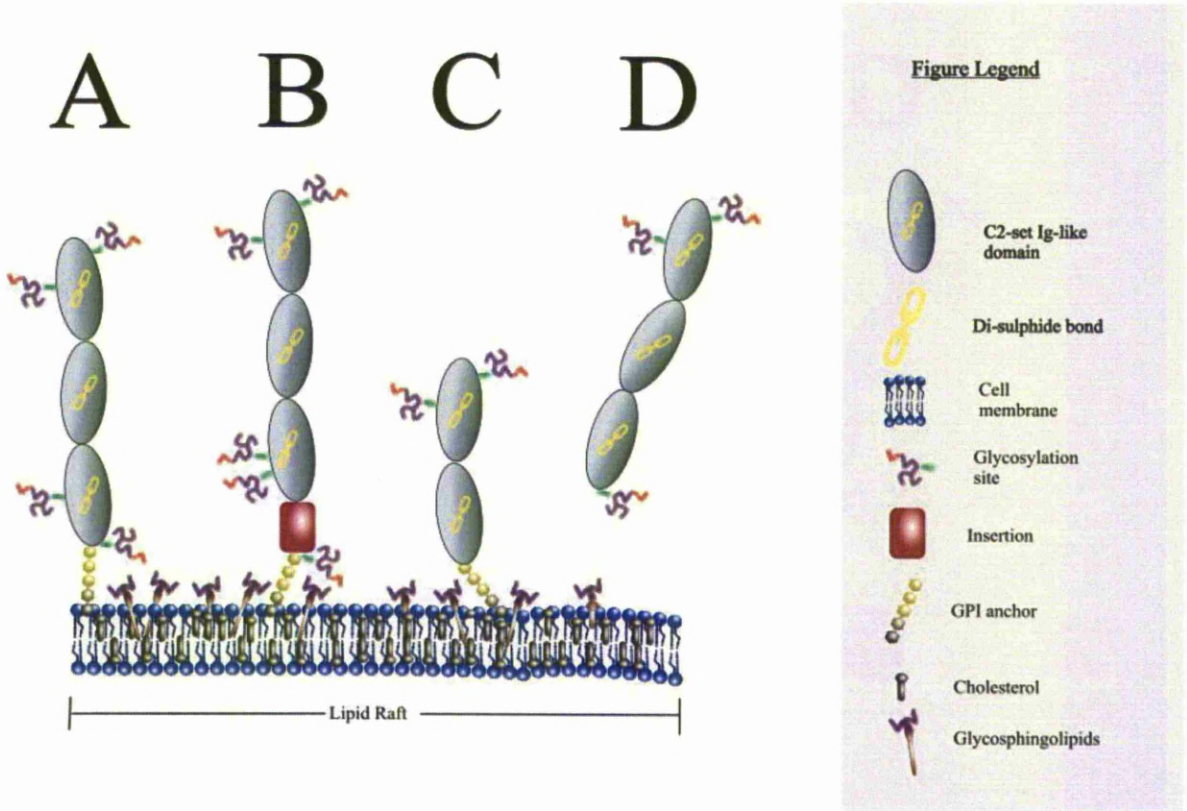


Figure 9 IgLON cell adhesion molecules. Most IgLONs have 3 Ig-like domains, are glycosylated, GPI anchored and localised to micro-membrane regions of the cell membrane. These micro-membrane regions are enriched with cholesterol and sphingolipids and are known as lipid rafts. (A) αLAMP, OBCAM, αCEPU-1 and the long form of NTR(NTR-L) typifies this isoform. However (B) β isoforms of *LSAMP* and CEPU-1 have been characterised with an insertion between the GPI anchor and its nearest Ig-like domain. (C) A short form of NTR (NTR-S) has been isolated that lacks the third Ig-like domain and (D) in chicken, a form of CEPU (CEPU-1-se) lacking the GPI anchor and is therefore soluble Lodge 2001. Figure adapted from Mark Howard.

1.15.3 IgLON expression pattern

OBCAM, in rat, was found mainly in the cerebral cortex and hippocampus but at low levels in the cerebellum, spinal cord and medulla oblongata whereas Kilon had a much more widespread expression pattern being located on the cerebral cortex, hippocampus, cerebellum, olfactory bud and diencephalon (Miyata et al., 2003a). OBCAM is synthesized within the somata, attached to vasopressin neurosecretory granules by the GPI anchor, and transported to the dendrites (Miyata et al., 2003b).

OBCAM was found as both mRNA and protein in the cat primary visual cortex at various postnatal ages. OBCAM expression level was high in young and diminished in older visual cortexes. OBCAM was also found in the soma and dendrites of pyramidal neurons in all cortical layers of the cats. Immuno-staining showed fewer OBCAM-positive neurons and neurites and at a lower staining intensity than younger cats however the OBCAM levels were higher in dark-reared cats than in those normally reared, at the same age, thus indicating that OBCAM may have a role in visual cortex development and plasticity (Li et al., 2006). Miyata's group furthermore showed that OBCAM was also located on cortical astrocytes and appears to be involved in control of cell proliferation and cell growth (Sugimoto et al., 2010).

LAMP protein is found on neural surfaces on the somata and proximal dendrites of neurons and axons of the brain during early development however this changes during postnatal development when in adulthood LAMP is only found on postsynaptic sites of the limbic neurons (Horton and Levitt, 1988, Zacco et al., 1990). LAMP, but not CEPU-1, was found in most of the migrating crest cells of the anterior region of the neural tube (Kimura et al., 2001). LAMP is also found in the mid- and hindbrain regions of rat (Reinoso et al., 1996).

In chick total-brain assays short transcripts, shown as g11/g19 in Figure 14, of *LSAMP* isoforms were found to be expressed equally at all stages from E8 whilst a longer transcript,

shown as g9 in Figure 14, isoform did not reach full expression until after E10 and became the prominent isoform at later stages (Brummendorf et al., 1997).

LAMP is found on the cortical and sub-cortical regions of the limbic-system e.g. amygdala, hippocampus, perirhinal and cingulate cortexes and the striatum of both the developing and adult brain (Horton and Levitt, 1988, Levitt, 1984, Pimenta et al., 1996, Reinoso et al., 1996).

LAMP has also been located in bone (Kresse et al., 2009), in cardiac smooth muscle cells (Wang et al., 2008) and indeed if the data cited in the BioGPS database (Wu et al., 2009) are correct LAMP is found in 76 tissues and biological compartments.

1.16 IgLONs bind homo- and heterophilically in *trans* and also bind in *cis.*

IgLONs bind both homo- and heterophilically in *trans* and there is also evidence of a *cis* binding interaction. Binding assays of homophilic FC recombinant IgLON proteins to IgLON expressing CHO cells (Figure 10 1A-1F) showed that *in trans LSAMP* had a higher affinity with Ntm/CEPU-1 and OBCAM and had a low homophilic interaction (Reed et al., 2004) although *LSAMP* had previously been shown to bind homophilically in trans (Zhukareva and Levitt, 1995, Pimenta et al., 1995). Embryonic, limbic system neurons exhibited strong neurite outgrowth on *LSAMP*-expressing CHO cells (Pimenta et al., 1995). It was expected that when CEPU-1 and OBCAM were co-expressed on the surface of CHO cells the binding to *LSAMP*-FC recombinant protein would be increased (Figure 10 2A) but, unexpectedly the binding force was reduced (Figure 10 2B) suggesting that Ntm/CEPU-1 and OBCAM were potentially binding in *cis* and in so doing the binding site for *LSAMP* was no longer exposed (Reed et al., 2004). This was the first evidence of a dimeric IgLON (DIgLON), formed by *cis* interaction of IgLONs within the plane of the membrane. This finding was later confirmed and expanded

on (McNamee et al., 2011) who showed that heterodynamic complexes form in the plane of the membrane.

The presence of two IgLONs either co-expressed on CHO cells or combined as a recombinant protein on an Fc tail, was found to have a functional effect. Outgrowth assays of E 7/8 chick forebrain neurons on doubly transfected CHO cells showed that neurite outgrowth initiation was inhibited by a substrate of DIgLON:CEPU-1-OBCAM (Figure 10 3A)and also DIgLON:CEPU-1-*LSAMP* (Figure 10 3B). Similarly, E8 neurite outgrowth was inhibited when DIgLON:CEPU-1-OBCAM-Fc (Figure 10 3C) recombinant protein was used as a substrate on poly-L-lysine base. E6 forebrain neurons, on the other hand, were unresponsive to the DiGLONs in the substrate, possibly owing to differences due to the stage of development, although analysis of the IgLON population expressed on the neurons appeared to show only marginal difference to those at E8 (McNamee et al., 2011).

LAMP has been shown to increase length of axons of *LSAMP*-expressing neurons within the limbic system (McNamee et al., 2011).

Neurons phosphatidylinositol-specific phospholipase- C-stripped of their IgLONs, and collaterally all GPI-anchored proteins, were unresponsive to the DIgLON:CEPU-1-OBCAM-Fc recombinant protein suggesting that cell-surface IgLONs are required to respond to the signal (Akeel et al., 2011).

As IgLONs are GPI anchored to the cell membrane and therefore having no transmembrane component it is not known how they signal to the cell's interior. Because pertussis toxin was shown to negate the DIgLON:CEPU-1-OBCAM-Fc-generated inhibition, it is possible that part of the signal transduction pathway is facilitated by a G-protein coupled receptor (Akeel et al., 2011; Clarke and Moss, 1994, 1997). IgLON binding interactions are probably affected by

the mix of IgLONs expressed. Although homo-dimeric *LSAMP*-Fc recombinant protein had been shown to bind strongly to CEPU1 and to OBCAM expressed on the surface of CHO cells the lower than expected binding affinity between DIgLON:CEPU-1-OBCAM and *LSAMP*-FC (Reed et al., 2004) hinted at complexities in how IgLONs might bind as the mix in the population of IgLONs changed. For example, increased expression of CEPU-1 could result in a scenario where the number OBCAM homo-dimer DIgLONs, with their high *trans* affinity for *LSAMP*, changes to a population of CEPU-1-OBCAM DIgLONs, with their low *trans* binding affinity resulting in greatly reduced adhesion and increased cell motility. In certain circumstances this may be unwanted and contribute to metastasizing of malignant cells.

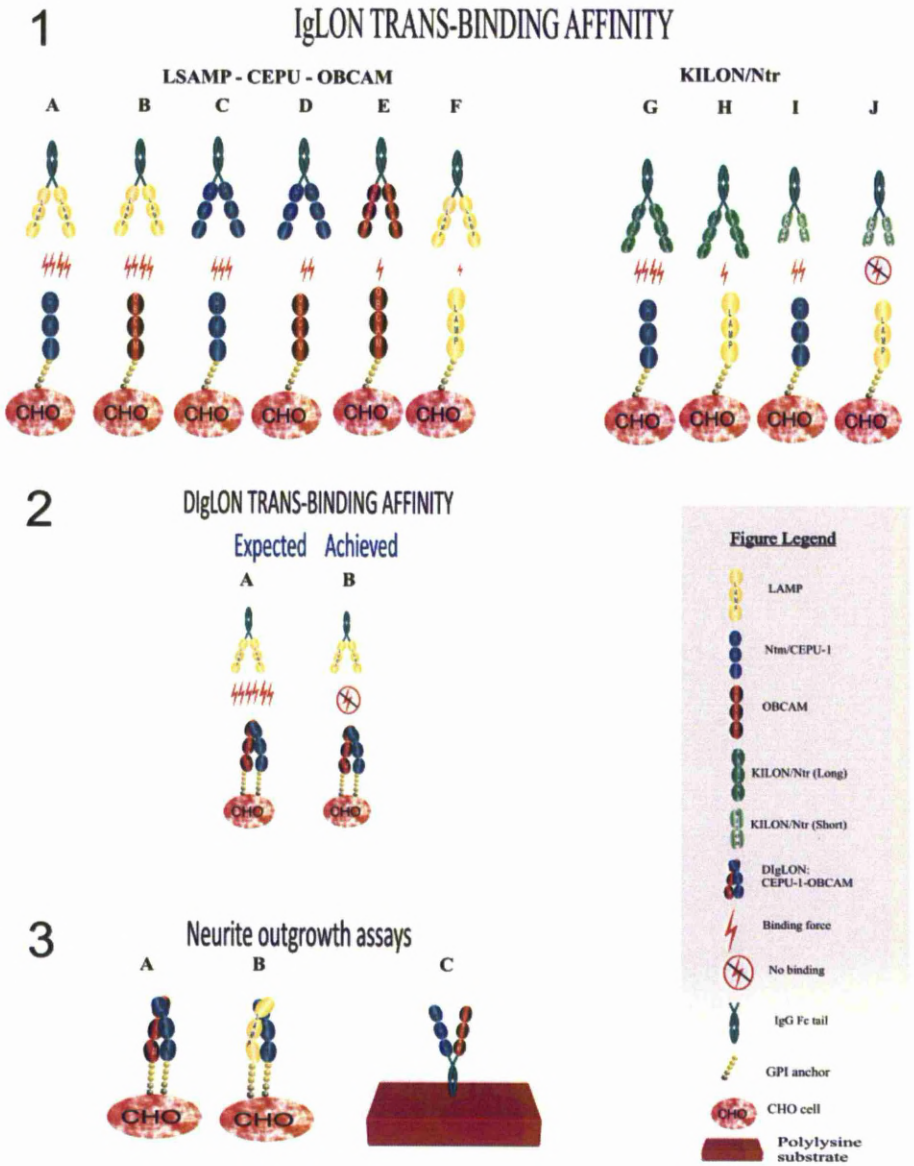


Figure 10 IgLON and DIgLON interactions. 1 (A-F) Hierarchy of trans-binding affinities exhibited by recombinant homodimers attached to Fc tails. (A,B) *LSAMP*-Fc binds strongest to CEPU1 and OBCAM with binding affinity decreasing in (C) CEPU1-Fc↔CEPU1 (D) CEPU1-Fc↔OBCAM (E) OBCAM-Fc ↔ OBCAM and (F) *LSAMP*-Fc ↔*LSAMP*.

1 (G-J) Binding hierarchy of KILON/Ntr (Long) with (G) Ntm/CEPU-1 (H) *LSAMP* Binding hierarchy of KILON/Ntr (Short) with (I) Ntm/CEPU-1 [S form lower binding force than with L form] (J) *LSAMP* [no binding].

2 It was expected that (2A) LAMP-Fc would have a very high binding affinity with DIgLON:CEPU-1-OBCAM the contrary was found to be the case with (2B) *LSAMP*-Fc having very low to no binding affinity with DIgLON:CEPU-1-OBCAM

3 . In neurite outgrowth assays (3A) DIgLON:CEPU-1-OBCAM, (3B) DIgLON:CEPU-1-*LSAMP* and recombinant protein (3C) CEPU-1-OBCAM-Fc all inhibited neurite outgrowth. (Reed et al., 2004) (Akeel et al., 2011)

1.16.1 IgLON expression pattern and binding in chick

In chick, LAMP expression was found to be already active at HH stage 10 in the anterior-most region of the neural tube However at HH stage 14, the expression was limited to most of the neural crest cells, the anterior midbrain region and in the notochord. The notochord expression was absent at HH stage 25 at which stage expression had moved to the floor plate and the lumbar level of motoneuron columns (Figure 11). At the later stages of development, expression in motoneuron columns was seen at all levels of the spinal cord, but these data were not shown(Kimura et al., 2001)

Figure 11 Localisation of LAMP mRNA in chick embryos.
Whole-mount *in situ* hybridization (A) lateral view of HH stage 14. Small triangles indicate expression in the neural crest. Large triangles indicate expression in the notochord. Arrows indicate expression in the anterior midbrain. (B) Transverse section at the lumbar levels of chick embryos at HH stage 25. White arrows indicate expression in the floor plate and asterisks indicate the expression in motoneuron columns.
(Kimura et al., 2001)

Interestingly, the expression pattern and binding ability differ between some mammalian and chick IgLONs. For example, in chick the KILON/Ntr homophilic interaction identified on Ntm-expressing CHO cells (Gil et al., 1998) was absent (Marg et al., 1999).

Ntm/CEPU1 forms noncovalent homodimers and multimers at the cell surface as evidenced by cross-linking studies of transfected CHO cells. Reaggregation of the transfected CHO cells

and the specific binding of an Ntm/CEPU1-Fc construct to the transfected CHO show that homophilic adhesion in *trans* is mediated by Ntm/CEPU1.

Differential adhesion force between the different IgLON isomers' was also described (Marg et al., 1999). In *trans*, the three-Ig-domain form of neurotractin (neurotractin –L) bound strongly to CEPU-1 (Figure 10 1G) but weakly to LAMP (Figure 10 1H). No homophilic binding to neurotractin-L was observed (Figure 10 1I). The two-Ig-domain form of neurotractin (neurotractin-S) bound less strongly to CEPU-1 (Figure 10J) than the L form while no LAMP binding could be detected (Figure 10K).

1.17 Transient adhesion and locomotion

Transient adhesions may be formed by non-junctional contact where the plasma membranes run parallel, typically with a 10-20nm gap between them. This non-junctional contact may be optimal for cell locomotion - close enough for traction and to allow transmembrane proteins to interact but not sufficiently close as to anchor the cell. Transient adhesion may also be involved in positioning and orienting the plasma membranes prior to the structural junctions being formed.

1.17.1 Can LSAMP bridge the intracellular gap?

A single Ig domain of ICAM-1 has been measured at being between 3.3nm and 3.7nm (Staunton et al., 1990) whilst NCAM and VCAM-1 are between 3.5nm and 3.7nm (Becker et al., 1989, Osborn et al., 1994). Some IgCAMs protrude extracellularly by quite a distance (Figure 7 Figure 8) and it easily imagined how they can physically interact. For example it is known that NCAMs bind NCAMs intracellularly. NCAM 140 consists of 4 Ig-like domains and 2 Fibronectin III domains. The Ig-like domains alone would allow each NCAM 140 to protrude

14nm from the plane of the membrane allowing two molecules on opposing cell surfaces to easily bridge the 10-20nm intercellular gap.

The LAMP protein is shorter than NCAM consisting of only three Ig-like domains, however this would allow LAMP protrude from the plane of the membrane by about 10nm therefore intercellular transient binding of LAMP is theoretically feasible although it is proposed that on forebrain neurons at least, IgLONs interact with an, as yet unidentified, IgLON-containing receptor complex (Akeel et al., 2011) as it was shown that DIgLONs inhibit the initiation of neurite outgrowth from forebrain neurons. Because the CAMs involved in this transient adhesion do not necessarily require firm anchorage or transmembrane signalling regions they may be one of, or act like, the non-classic cadherins, such as the truncated, GPI-anchored T-cadherin (Ranscht and Dours-Zimmermann, 1991). GPI-anchored surface proteins having no cytoplasmic tail have no direct intracellular communication mechanism nor are they connected to the cytoskeleton although, in a poorly understood way, T-cadherin, for example, facilitates calcium dependent homophilic binding *in vitro* (Vestal and Ranscht, 1992). Another signalling mechanism for CAMs lacking a transmembrane moiety that has been suggested is that the stress and change in shape of the cell due to adhesion may act as a signal (Basson, 2008). GPI-anchoring results in a weak junction that probably relies on the Velcro principal of many weak adhesions providing a large overall result (Alberts et al., 2008). This weak anchoring could also be provided by certain members of the immunoglobulin-like-domain-containing cell adhesion molecules (IgCAMs) family.

1.17.2 Synaptic Formation is highly active during development and continues at a lower rate throughout life.

Synaptogenesis describes the formation of synapses. Synapses allow, typically, a chemical signal to be passed from a neuron, via its axon, to its target. Synaptogenesis is particularly

active during neural system development and continues at a lower rate throughout life. The current model of synaptogenesis envisages three phases; attraction, contact and stabilization. Attraction of the axon to the general target area is facilitated by the repulsive and attractive substrate- and target-derived cues (Tessier-Lavigne and Goodman, 1996). Upon arrival in the target area more precise steering of the axon to its target may be required (Kalinovsky et al., 2011) eventually leading to initial contact between the axon and its specific target. This initial contact is typically made by filopodia that extend from both the axons and dendrites (Ziv and Smith, 1996, Jontes and Smith, 2000). Finally, via pre- and post-synaptic signalling, the desirable connections are identified and stabilized whilst the undesirable ones will fail to develop (Goodman and Shatz, 1993, Katz and Shatz, 1996).

Some IgSF proteins have been implicated in synaptic formation and interactions. NCAM functions as a homophilic CAM and, for example, is found on surface membranes of the mossy fiber terminals where they contact the apical dendrites of pyramidal neurons, and is also found, in rat, at both the pre- and post-synaptic membranes of some spine synapses in the dentate gyrus (Schuster et al., 2001). SynCAM overexpression has been implicated in either increasing the number of synaptic terminals or enhancing pre-synaptic neurotransmitter release (Biederer et al., 2002).

IgLONs are also involved in synaptogenesis; a single and double overexpression analyses indicated that IgLONs are involved in the regulation of synapse formation in hippocampal neurons. The single overexpression of LAMP or OBCAM increased dendritic numbers whereas that of CEPU/Kilon reduced dendritic numbers and Ntm was neutral in effect. Interestingly, neither the double overexpression of LAMP nor OBCAM had any effect on dendritic formation. When neurons overexpressed Kilon-Neurotrimin, LAMP-Neurotrimin, OBCAM-Neurotrimin and OBCAM-Kilon a significant reduction in dendritic growth was observed. The difference in the observed results from the homophilic and heterophilic

overexpression indicates that it is possible that homophilic and heterophilic *cis*-interactions on the post-synaptic site interact with pre-synaptic IgLONs to modify synapse formation (Hashimoto et al., 2009). In mouse hippocampal neurons OBCAM was found to be localised to post-synaptic dendritic spines, with a reduction and increase in synapses on dendrites matching the inhibition and overexpression of OBCAM (Yamada et al., 2007). This indicates that OBCAM is a synaptic CAM involved in synaptogenesis.

1.18 IgLONs may have a role as tumour supressing genes.

IgLONs may act as antioncogenes since loss or reduction of *LSAMP* expression correlates with renal clear cell carcinomas (Chen et al., 2003). A reduction of OPCML expression was found in epithelial ovarian cancer (Sellar et al., 2003) and also in gliomas and other brain tumours (Reed et al., 2007). In epithelial ovarian cancer, reduced expression of *OPCML, LSAMP*, and *NEGR1*, was found in the majority of tumours, but, conversely, h*NT* expression was increased (Ntougkos et al., 2005). CEPU1/NTM appears to have a role in mediating oestrogen-induced sympathetic pruning in some peripheral targets (Krizsan-Agbas et al., 2008).

1.19 *LSAMP* is a large and complex gene.

The *LSAMP* gene, in chick, has three isoforms and, depending on the isoform, consists of seven or eight exons (Table 5). It has two exons 1, named exons 1a and 1b, containing the distal and proximal transcription start sites (TSSs) respectively from which the 3 transcripts are generated, a long form of 970 kbp and two shorter ones of ~286 kbp (Figure 14). These transcripts each code for a separate protein isoform, the long transcript codes for the g9 isoform and the short transcript codes for the g11 and g19 isoforms (Figure 13, Figure 14Figure 14). The LAMP protein isoforms (g11/g19 in chick) coded for by the short

transcripts differ in that g11 has an amino acid insertions towards the 3' end consisting of 23 amino acids (human, rat, mouse (Pimenta and Levitt, 2004)) and in chick, 12 amino acids (Figure 13). The function of this insertion is unknown but it hypothesised that extra flexibility and length conferred on the protein would be an advantage for molecular interactions (Brummendorf et al., 1997). Although Brummendorf calculated that the g9 and g11 would start and end at the same location this is not supported by the Ensembl entries (Figure 14) which shows different start and end locations for all three isoforms.

The *LSAMP* gene in human extends for greater than 2.1 Mbp and consists of 10 exons (Figure 12 and Figure 16). Like chick, the human form also has two exons 1 (Exon 1a and Exon 1b), containing alternative TSSs, that are separated by > 1.5 Mbp which contributes greatly to the large size. Three transcripts are generated by the gene, one from Exon 1a and two beginning at Exon 1b.

Surprisingly the long form of the human *LSAMP* gene is not documented in the nucleotide databases but was identified by cross species comparison using the Ensembl (Flicek et al., 2011) chick transcript ENSGALT00000024337 (Figure 15). This long extent spans 2.1 Mbp in human (970 kbp in chick).

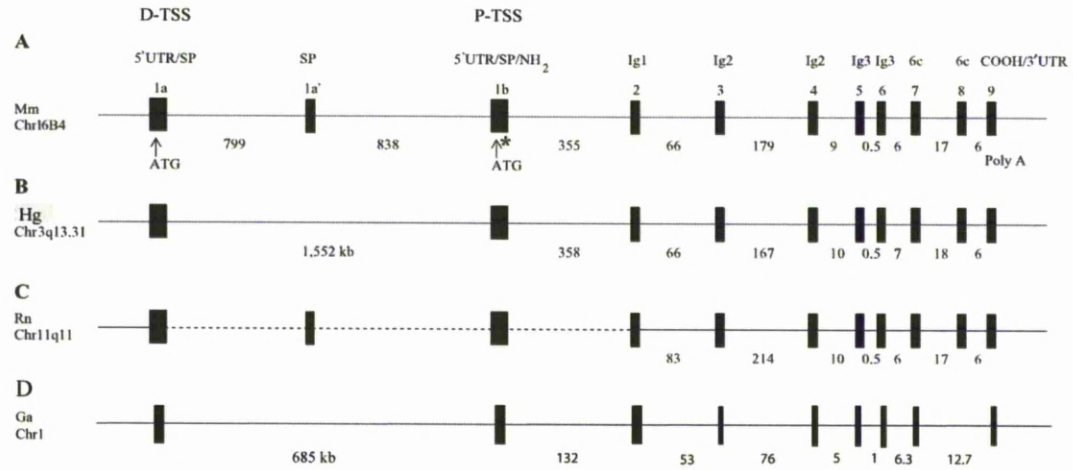


Figure 12 Comparison of the intron-exon structures between *LSAMP* orthologs in mouse and human. Genomic diagram of the *LSAMP* gene in (A) mouse (B) human (C) Rat (Pimenta and Levitt, 2004) and (D) Chicken (author's diagram, data taken from Ensembl (Flicek et al., 2011). The black oblongs indicate exons and the thin lines represent the introns with their genomic extent indicated below them, those in (D) may not be to same scale as A,B or C. The distances been exons are in kbp. The exons containing the upstream, distal TSS (marked D-TSS) and proximal TSS (marked P-TSS) are shown. The TSSs are arrowed in (A) mouse and indicated with ATG. The site where the distal TSS is spliced into the proximal TSS is indicated by *.

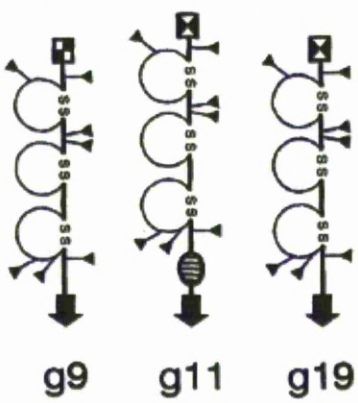


Figure 13 Domain models of chicken *LSAMP*-isoforms. g9 is the isoform with the longest genomic extent whilst g11 and g19 share the same TSS. Key: Immunoglobulin-related domains are drawn as loops which are closed by disulphide bridges. Putative N-linked glycosylation sites are shown as lines ending with triangles and the GPI-anchor is marked by an arrow. The alternatively spliced signal peptides are represented by different boxes and the internal alternatively spliced 12 amino acid segment is indicated by a hatched ellipse. (Brummendorf et al., 1997)

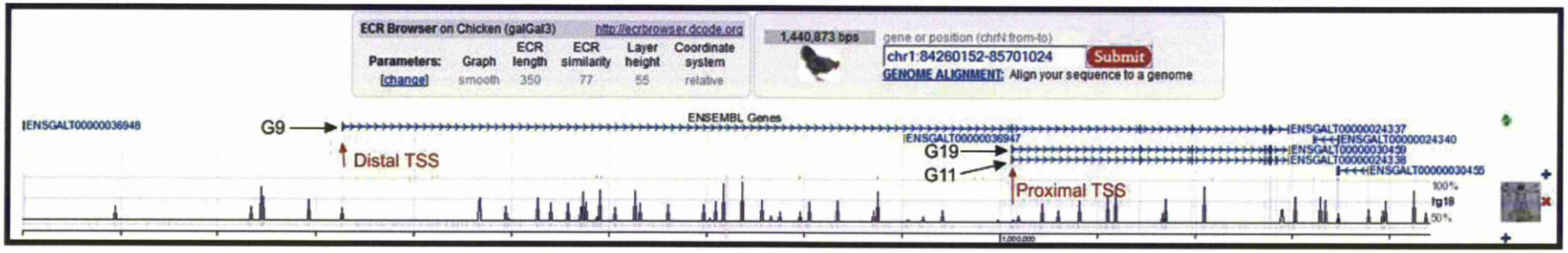


Figure 14 Genomic view of *LSAMP* gene (chicken as base organism, human as target) showing the three Ensembl transcripts for *LSAMP* ENSGALT00000024337, ENSGALT00000030458 and ENSGALT00000024338 correlating to Brummendorf's g9, g11 and g19 respectively (Brummendorf et al., 1997). The distal and proximal TSSs are indicated.

Brummendorf	UniProt	Ensembl (Gene name: ENSGALG00000015087)						
Name		Name	Location	Genomic extent (kbp)	Transcript ID	Protein ID	Length (aa)	Exons
G9	SN = ChLAMP, g9-isoform DBID = O02870	O02870_CHICK	Chr 1 84,587,331-85,557,493	970.16	ENSGALT00000024337	ENSGALP00000024291	334	8
G11	SN = ChLAMP, g11-isoform DBID = O02869	O02869_CHICK	Chr 1 85,272,074-85,557,507	285.43	ENSGALT00000024338	ENSGALP00000024292	350	8
G19	RN = Limbic system-associated membrane protein AN = CHLAMP G19-isoform AN = E19S DBID = Q98919	*LSAMP*_CHICK	Chr 1 85,271,704-85,558,083	286.38	ENSGALT00000030459	ENSGALP00000029822	338	7

Table 6 Nomenclature and data for *LSAMP* isoforms from Brummendorf (Brummendorf et al., 1997), UniProt (Consortium, 2012) and Ensembl (Flicek et al., 2011). Key: RN = Recommended name; AN = Alternative Name; SN =Submitted Name; DBID = UniProt Knowledge Base Database identifier;

1.19.1 LSAMP's Genomic neighbourhood – an overview.

Although in human, *LSAMP* is on chromosome 1 and in chick it is on chromosome 3 there are similarities in the genomic vicinity. In both cases *LSAMP* is flanked by the same genes although they have different names depending on the species and the database (Figure 15 and Figure 16). The nearest protein-coding characterised genes are Immunoglobulin Superfamily Member 11 (IGSF11) and Growth-associated protein 43 (GAP43). It is interesting to note that these three genes are related by family and/or function. LAMP, as has been discussed, is a member of the immunoglobulin –superfamily (IgSF) and is a cell adhesion molecule, IGSF11 is also a member of this family and functions as a cell adhesion molecule that through homophilic interaction stimulates cell growth (Harada et al., 2005). Neuromodulin, the protein encoded for by *GAP43*, is expressed at high levels in neuronal growth cones during development and axonal regeneration (Schleinitz et al., 2008). These similarities might indicate that that these genes may have originated from mutations to a common ancestral gene. There is however little similarity between the resulting peptide sequences coded for by the three genes (Table 6). It is also feasible that these genes may form part of an expression cluster whereby energy efficiency is achieved by the co-localisation of genes that are concurrently expressed. Although there may be an energy advantage a recent study has shown that if a cluster is perturbed transcription rates do not appear to be affected (Britton et al., 2011).

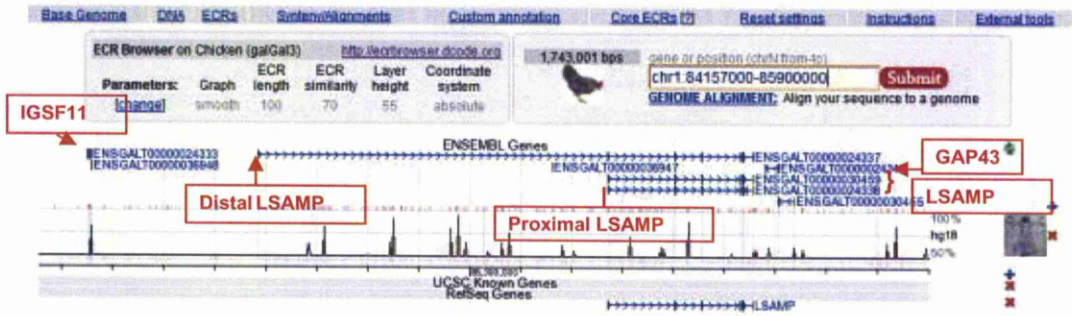


Figure 15 ECR Browser (Ovcharenko et al., 2004) (using Ensembl nomenclature) compared using chick as base organism and human as target organism showing long form of *LSAMP* (ENSGALT00000024337), with its TSS identified by "Distal *LSAMP*", and the two shorter forms marked as *LSAMP* Short and indicated with '}'. Their TSS is marked "Proximal *LSAMP* TSS". The nearest characterised genes IGSF11 and GAP43 are also indicated.

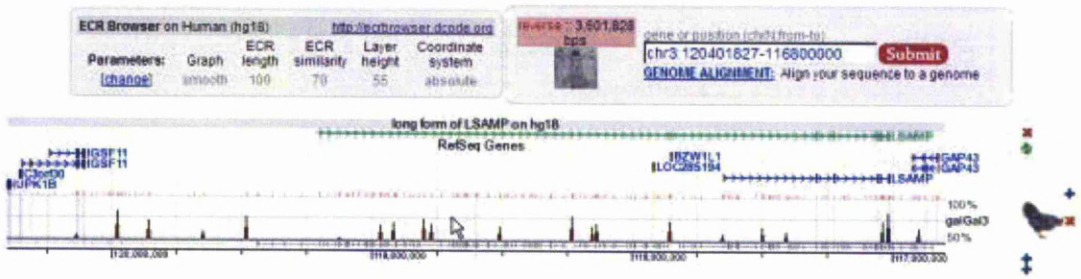


Figure 16 Genomic region surrounding *LSAMP* in human. ECR Browser output with human as base (reversed orientation to match chick) and chicken as target (Stringency 100bp/70% similarity) with the region downstream of *LSAMP* to its nearest neighbouring gene, IGSF11. One of the short human transcripts of *LSAMP* (RefSeq) is shown in blue; the second short transcript was not included in the human REFSEQ database but is similar in length and position to the annotated, short version. The long human transcript is shown in green and is manually inserted.

	IGSF11_IsoB (431aa)	IGSF11_IsoA (430aa)	Neuromodulin_Iso1 (274aa)	Neuromodulin_Iso2 (238aa)	*LSAMP*_338aa (338aa)	*LSAMP*_345aa (345aa)	*LSAMP*_205aa (205aa)
IGSF11_IsoB		96	6	6	15	14	16
IGSF11_IsoA			6	6	15	13	16
Neuromodulin_Iso1				95	6	6	3
Neuromodulin_Iso2					7	7	3
*LSAMP*_338aa						94	71
*LSAMP*_345aa							71

Table 7 Comparison of *LSAMP*, IGSF11 and Neuromodulin using ClustalW (Larkin et al., 2007). Little correlation was found between the different proteins although the different isoforms of the same proteins naturally did show high similarity.

Isoforms A and B of IGSF11 (NP_689751.2 ,NP_001015887.1),

Isoforms 1 and 2 of Neuromodulin/GAP43 (NP_001123536.1, NP_002036.1) NCBI protein sequences and *LSAMP* (NP_002329.2 (NCBI) or ENSP00000328455 (Ensembl) at 338 a.a. and also Ensembl sequences *LSAMP* ENSP00000419000 (345 a.a). ENSP00000418506 (205 a.a.).

KEY: SeqA/SeqB indicate the target and object respectively during the comparison process, length is the amino acid length of the protein sequence and the score indicates similarity, a higher score indicates greater similarity.

1.19.2 LSAMP is inserted in human in the opposite direction to that in chick

Over the course of evolution a chromosomal inversion of the region containing *LSAMP* appears to have occurred in various species. This has resulted in *LSAMP* being inserted in, for example, rat and human in reverse orientation (qter to pter) as is *IGSF11*, whereas *GAP43* is inserted in the forward orientation (pter to qter). All these genes are in the forward orientation in mouse and chick.

1.19.3 Small and pseudo genes are also present in LSAMP's genomic neighbourhood.

There are other smaller genes and/or pseudo genes also in this genomic region. Two of these are in the intronic region within the *LSAMP* transcription unit and reside between the two TSSs. These genes are the non-protein-coding RNA gene, LOC285194 (forward orientation) and the pseudogene BZW1L1 (reverse orientation).

Human gene *LSAMP*-AS1 (aka *LSAMP*-AS3, Ensembl gene ENSG00000243197, Entrez Gene: 285194) on the positive strand, produces six non-coding RNA transcripts, *LSAMP*-AS3-001 through *LSAMP*-AS3-006 (Figure 17). The short form of human-*LSAMP* is located from 117,011,832-117,647,068, in reverse orientation on chromosome 3, the *LSAMP*-AS-00x pseudo genes are located downstream from *LSAMP* at 116,428,626-116,442,428, with GAP43 being inserted between the active and pseudo genes.

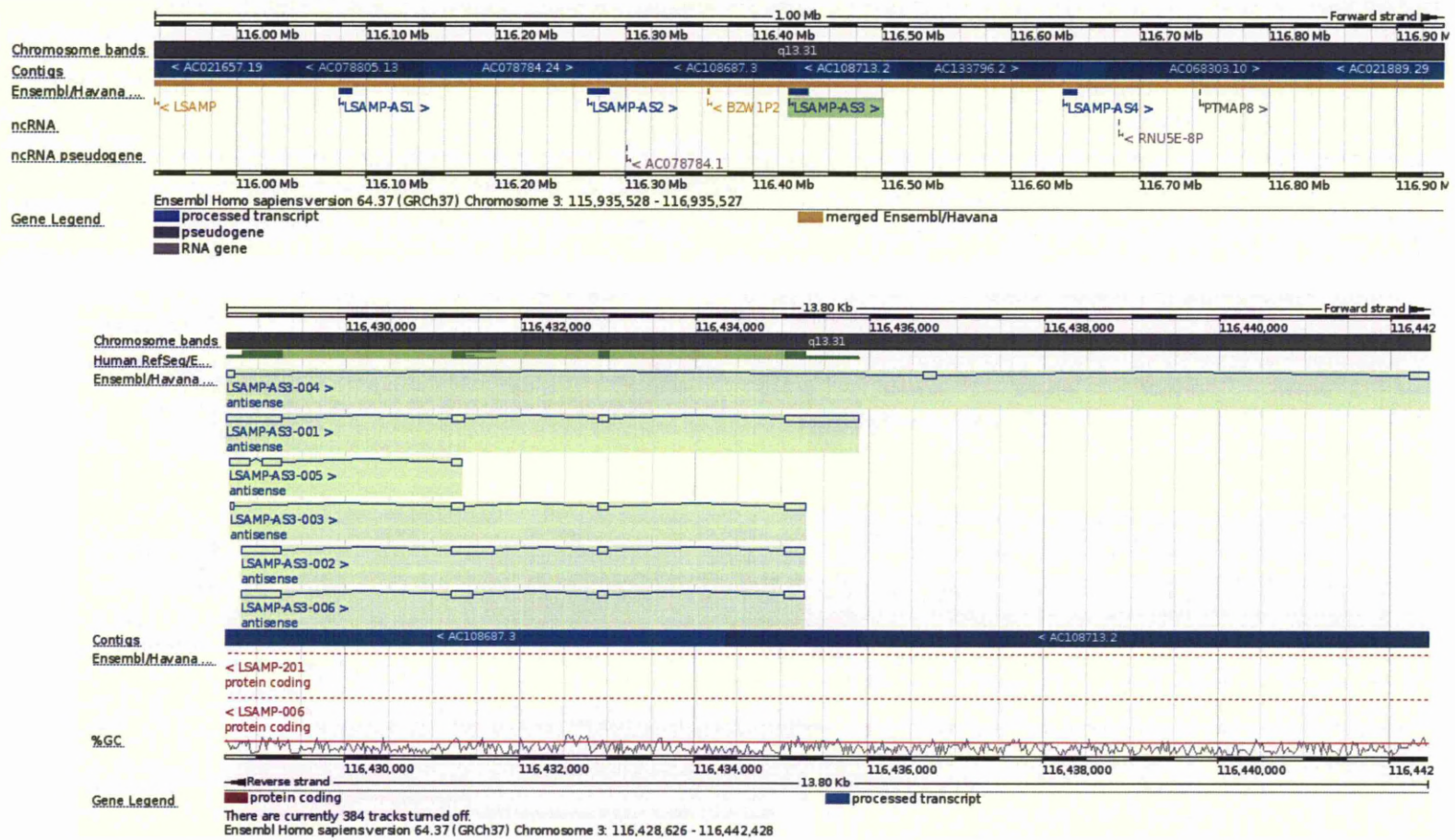


Figure 17 Ensembl's diagrams of *LSAMP* anti-sense non-coding RNAs. These are located 569 kbp downstream of *LSAMP* on chromosome 3.

1.19.4 The density of known genes varies between chromosomes

Some chromosomes such as the human chromosome 19 are gene rich (~25 genes/Mbp) whilst others such as human chromosomes 4 or 18 (~4 genes/Mbp) are more sparsely populated. Typically the genes on gene-dense chromosomes are smaller than those on the less dense chromosomes. The median size for genes on human chromosome 19 are ~12 kbp whilst on chromosome 4 they are ~38kpbs and chromosome 18 the median gene size is in the order of 50 kbp. For example, in human, the *Hb9* gene is located on chromosome 7 and *LSAMP* is located on chromosome 3, both chromosomes have similar gene densities. Chromosome 7 is slightly more gene-dense at 6.96 genes/Mbp to chromosome 3's 6.08 genes/Mbp (Treit et al., 2010) . This, albeit only slightly lower gene-density, could allow for more control regions overall in the gene.

Although enhancers may be upstream, downstream, or intronic and are not confined to the immediate location of the gene, and don't even have to be on the same chromosome (Bateman et al., 2012, Spilianakis et al., 2005) many are found upstream of the gene they control (Mendelson and Quinn, 1993, Nakano et al., 2005, Iyer et al., 2010) . However this is further complicated in *LSAMP* because of the large region between the two TSSs allowing for the possibility of ECRs associated with many other genes to be located there. The chromosomal inversion also has the potential to complicate matters further because if some enhancers moved with the inversion and others did not, enhancers that were adjacent at one point may now be several Mbp away.

1.19.5 Known enhancers in the human LSAMP genomic neighbourhood.

A DNA sequence, hs253, identified by the VISTA enhancer browser (Visel et al., 2007)has been found to be an enhancer. This region is located in an intragenic locus of the human gene ZBZTB20 and is 907122 bp 5' (GRCh37.p2 NCBI reference assembly) to GAP43 (neuromodulin isoform 2) and was found to drive expression in the somite region of embryonic day 11.5 mouse embryos. Two other sequences in the same genomic region were not found to drive expression at this stage of development.

1.20 Aim of the project

The aim of the project was to initiate the development a chick-based model to enable rapid identification of the enhancers of gene expression. The *LSAMP* gene was selected as the first gene to be investigated due to continuing interest within our laboratory of the family to which *LSAMP* belongs, and because of recent findings that indicate a possible link with anxiety and other affective disorders.

CHAPTER 2.

2 Materials and Methods

The protocols used in the molecular biology described in this thesis, for general work were from the Cold Spring Harbor manual (Sambrook et al., 1989) and online (2006) or were as supplied by the manufacturers with their kits.

Reagents used are detailed in Table 20 and equipment used is in Table 21 and solution recipes are in Table 22 in the Appendix. Methods that differ from the generic description here are detailed in the relevant chapters.

2.1 Reporter plasmids:

2.1.1 pCRII

The pCRII vector, supplied by Invitrogen as part of its TA Cloning kit, is used for the direct cloning of Taq-amplified PCR fragments. Commonly used Sp6 and T7 promoter regions are included to facilitate sequencing. The multiple cloning sequences (MCS) contains restriction sites some of which bracket the TA cloning site and other restriction sites that are unique and permit directional insertion in downstream applications. Blue/white screening is also supported due to the inclusion of the LacZ gene. pCRII contains ampicillin and kanamycin resistant genes to allow for selection in suitable medium.

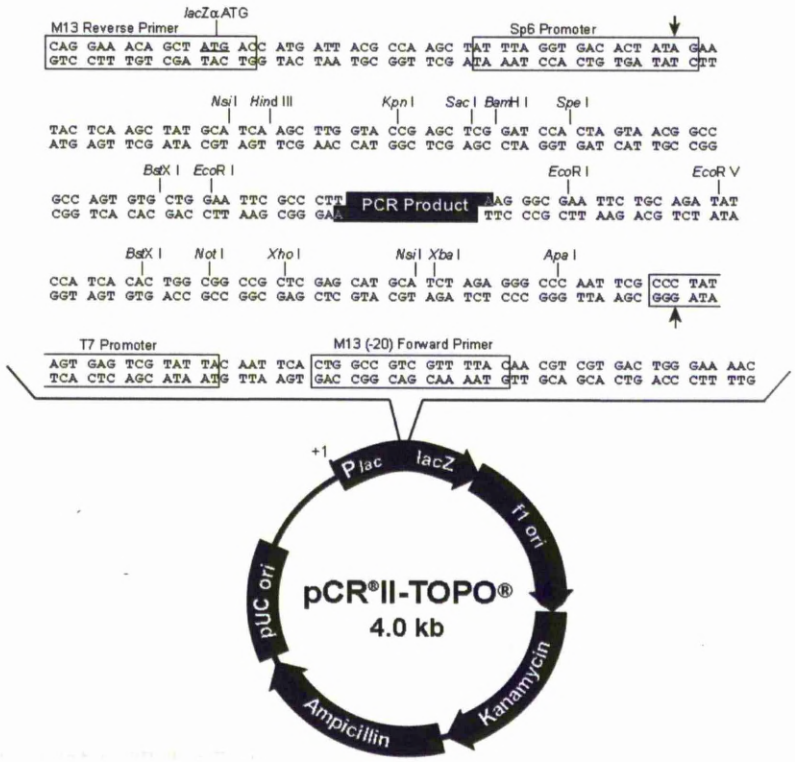


Figure 18 Map of cloning vector pCRII (Courtesy of Invitrogen)

2.1.2 p1229

The p1229 reporter vector allows for testing of enhancer functionality (Yee and Rigby, 1993). p1229 was obtained as a gift from Dr. Alasdair MacKenzie (University of Aberdeen) and is a modified version of pBLUESCRIPT, with the entire *LacZ* ORF/cDNA under the control of a human βglobin minimal promoter inserted into the polylinker area.

This vector requires that an enhancer sequence is cloned upstream of the promoter, for which purpose an MCS is provided. When transcription factors bind the cloned enhancer sequence the human βglobin minimal promoter is activated and the LacZ gene is transcribed. This results in the expression of β-Gal by the cell. The β-Gal is later visualised by staining.

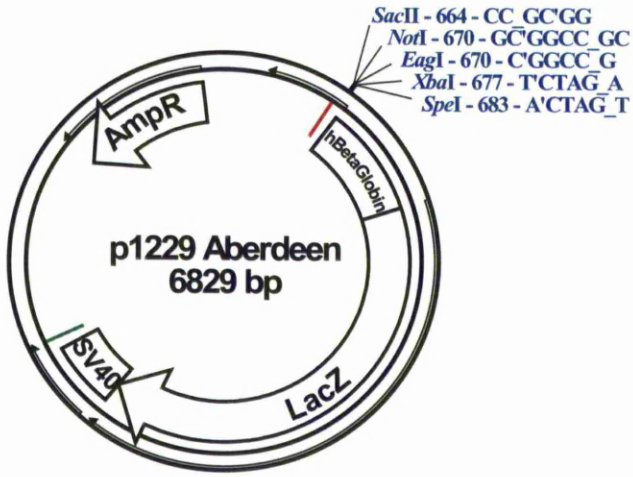


Figure 19 Map of p1229. The location of the Ampicillin resistance gene (AmpR) and minimal human β-globin Promoter (hβglobin), LacZ gene and SV40 (PolyA generation) insertion into the pBluescript backbone. The T7 sequencing promoter is shown marked with a red line and the T3 sequencing promoter with a green line. DNA sequence supplied by Alasdair MacKenzie. The RE sites at the MCS that cut only once in the whole sequence are shown. For this thesis the Evolutionary Conserved Regions (ECRs) were inserted between NotI (+670 bp) and SpeI (+683 bp). Arrows indicate ORFs. Map generated using pDRAW32 (pDRAW32 by AcaClone Software) from sequence supplied by Alasdair MacKenzie.

2.2 GFP Plasmid

In order to demonstrate that transfection had occurred and to locate this site a constitutively active GFP-expressing vector was co-injected with the p1229-based vector. The plasmid used was an IRES-containing GFP-expressing reporter plasmid pCaβlinkmigiresEGFPm5Cla1 - a gift from Jon Gilthorpe (Department of Developmental Neurobiology, King's College London). A β-actin promoter upstream of the internal ribosome entry site (IRES) under influence of the cytomegalovirus enhancer resulting in constitutive expression by the plasmid resulting in high levels of expression of the, downstream, enhanced green fluorescent protein (eGFP) gene (Yaneza et al., 2002). The GFP

signal was visualised by illumination via a band-pass filter centred at 488 nm using a fluorescent-capable microscope.

2.3 Molecular Biology Protocols

2.3.1 Construction of p1229 Reporter Vector Construct.

The transgenic reporter vector was constructed by PCR-amplification of a putative enhancer sequence and subsequently introducing it into the p1229 plasmid. This had been previously opened at the NotI/SpeI sites in its MCS. Following cloning and population expansion the transgenic p1229-based plasmids were extracted (Figure 20). Chick embryos were co-injected and transfected with the transgenic vector and also with the constitutively active IRES-GFP. The IRES-GFP was used as a marker of the injection site and a gauge of the efficacy of the transfection. Following incubation the embryos were stained to visualise the β-Gal expression.

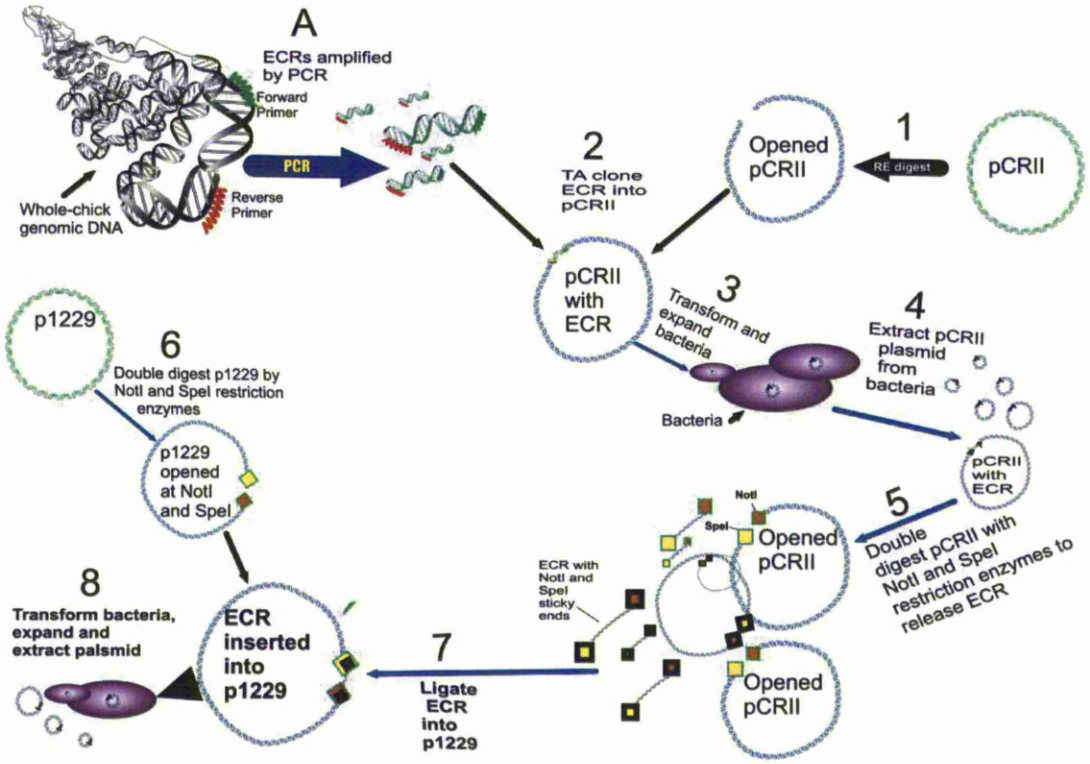


Figure 20 Amplification and Insertion of ECR into p1229. (A) PCR Expansion of ECR sequence from whole-chick genomic DNA. It is assumed that a single, 3'-adenine overhang will have been added to each end of the PCR product. (1) pCRII is supplied in the TA cloning kit, linearized and with a T overhang (2) The ECR fragment is cloned into the pCRII plasmid. (3) Competent bacteria (DH5α or TOP10F') are transformed with the pCRII:ECR plasmid construct and the population of bacteria expanded. (4) Using a plasmid-prep kit, the bacteria are lysed and plasmids are harvested. (5) pCRII:ECR plasmid construct is digested with NotI and SpeI to release ECR, now with NotI and SpeI sticky ends. (6) p1229 plasmid is opened at NotI and SpeI restriction sites and (7) p1229:ECR construct is formed by ligation. (8) Competent bacteria are transformed, population expanded and bacteria lysed to release the p1229:ECR construct.

2.4 Identifying putative enhancer sequences

2.4.1 ECRBrowser

An ECR is a region of non-coding DNA sequence that has been conserved across species over evolutionary time. The ECR Browser (Ovcharenko et al., 2004) gives a graphical representation of the regions conserved between the base species and target species. The region defined as being conserved is subject to the parameters of the length and percentage

similarity of the sequences. The ECR Browser's default is that a sequence with a length of ≥ 100bp and at least 70% identity between the sequences will be flagged as being an ECR. These criteria may be altered by the user.

Thirteen species are included in the database any of which may be selected as the 'base' against which one or all of the other organisms can be aligned and the conserved regions identified.

The ECR Browser key (Figure 21) is the example given on the website (http://ecrbrowser.decode.org/) to illustrate the features of the latest version of the ECR Browser's graphical representation by using the conservation between human and mouse of the SRPRB gene as an example.

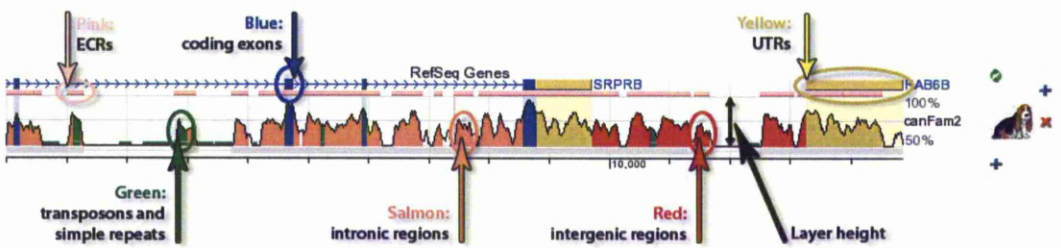


Figure 21 Screenshot from the ECR Browser site showing the key to the graphical representation. ECRs are shown as a pink track encompassing both coding and non-coding genomic regions. The coding exons (blue), intronic regions (salmon) intergenic regions (red), transposons and simple repeats (green) and untranslated regions – UTRs (yellow) are shown as coloured peaks on a graph; the x-axis represents positions in the base genome and the y-axis represents percentage identity between the base and target genomes.

.

The ECR Browser was used as follows:

The base organism was selected (e.g. mouse) and the gene name or genomic location was inserted in the location field. The request was submitted.

The list of available species was returned by the ECR Browser and the required species was/were selected (e.g. chick). The stringency of the selection was redefined as required to produce a well-defined region identified as being an ECR and the request submitted. In this case the default value of length 100bp/70% identity was retained (Figure 22). The sequence length and location represented on the screen was adjusted as appropriate using the zoom and horizontal scroll functions. Placing the cursor on the ECR line (arrowed) linked to the ECR sequence of both the base (mouse) and target (chick) organisms. This function also gives the option to extend the sequence 100bp each side of the ECR, this is convenient when designing primers. The sequence (Figure 23) was copied and used to design the primers.

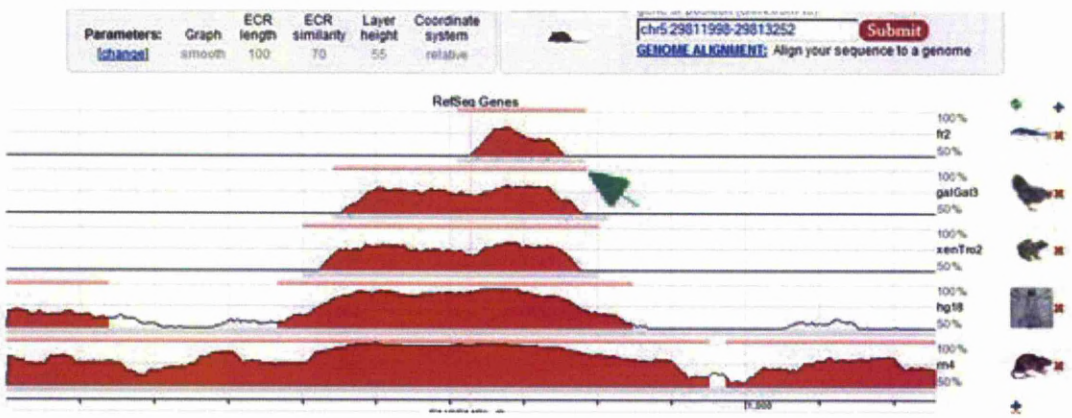


Figure 22 ECR Browser output showing cross-species homology with Mouse as the base organism and target organisms in order of similarity, rat, human, frog, chicken and puffer-fish. ECR indicated by green arrow. Stringency at default (100bp/70%).

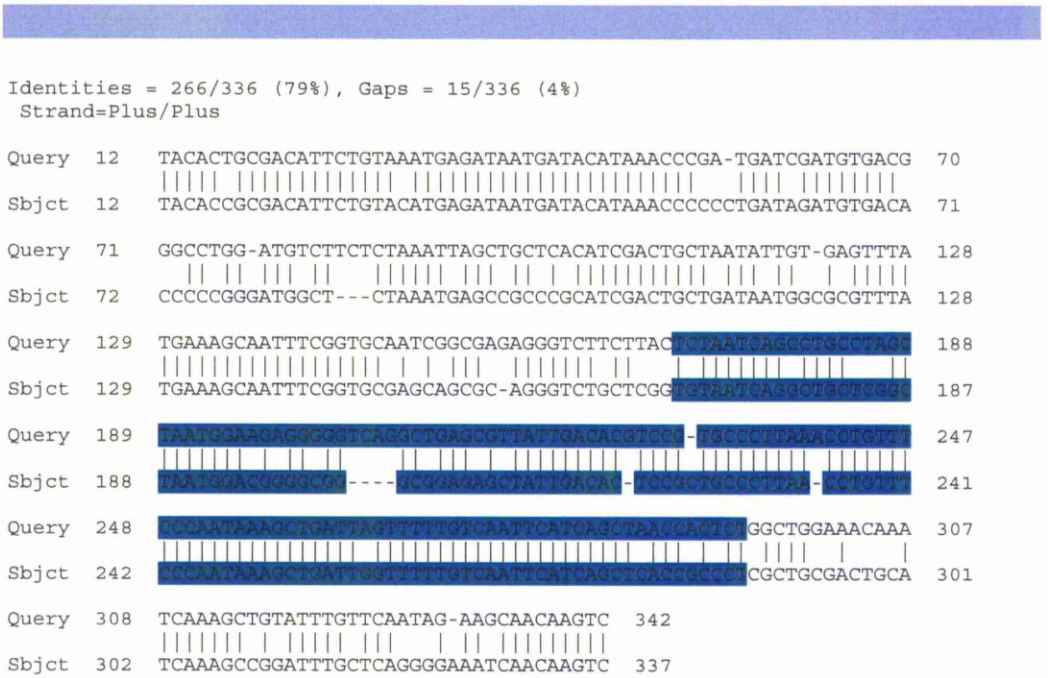


Figure 23 Details from the ECR Browser grab function (base organism, mouse; target organism, chick) of the Hb9 ECR showing the mouse and chick ECR sequences. The mouse 125bp sequence and the equivalent chick 120bp sequence highlighted. Aligned using BLASTN (Altschul et al., 1990)

2.5 Primer Design and synthesis

Suitable primer sequences were generated using Primer3 (Rozen and Skaletsky, 2000) (release 1.0b) using default settings with the putative-enhancer sequences marked as source sequence with square brackets, thereby forcing the primers to flank the core sequence.

The primers were synthesized (Sigma-Aldrich) and the region was amplified by either standard or gradient PCR using *Taq* polymerase with 2.5ng/µl whole chick genomic DNA as the template.

2.6 PCR

Taq polymerase from different manufacturers was used; this depended on availability, being mainly supplied by Invitrogen and New England Laboratories. The mix for the PCR reaction was as per the manufacturer's instructions and consisted of the template DNA (e.g. the whole chick genomic DNA (Merck-Novagen cat: 69233-3) or plasmid with insert), the forward and reverse primers, *Taq* polymerase, dNTPs PCR buffer, $MgCl_2$, and nuclease-free water. An adjunct of DMSO 5% v/v or BSA at 0.1ug/1ul was occasionally used in an attempt to improve the product yield.

Two PCR thermal cyclers were used, HYBAID, PCR Sprint001 and a gradient PCR system (MJ Research PTC-200 DNA Engine)

The programs used in the two systems are as detailed below (Table 8 Table 9) and typically 25µl or 50µl volume was used in each PCR reaction.

Components	µl	Final concentration
10X PCR buffer minus Mg	5	1X
10 mM dNTP mixture	1	0.2 mM
50 mM $MgCl_2$	1.5	1.5mM
Primer mix (10 µM each)	2.5	0.5 µM (each)
Template DNA	1-10[1]	
Taq DNA Polymerase or Platinum® Taq (5 U/µl)	0.25	1.25 units [2]
Autoclaved distilled water	to 50ul	

Table 8 Standard PCR mix (Adapted from Invitrogen Part no. 18067017.pps Rev. date: 2 Jan 2007)

Notes: [1] Whole chick gDNA used at 2ng/µl

[2] When amplifying from whole-chick genomic DNA 2U taq/50µl reaction used.

Step	Repeat	Action	Mins	°C
1		Activate Taq	2	94
2	30	Denature DNA	1	94
		Anneal primer (T_A typically 2°C below T_M)	1	50 - 70
		Extend DNA	1	72
3		Final incubation	10	72
4		Hold	Infinity	4

Table 9 Standard PCR - Program 1/Step

Step	Repeat	Action	Time	°C
1		Activate Taq	10m	95
2	35	Denature DNA	20s	94
		Anneal primer	30s	50 - 70
		Extend DNA	40s	72
3		Final incubation	10m	72
4		Hold	Indefinite	4

Table 10 Gradient PCR - Program AAA

2.7 PCR Amplification of DNA.

2.7.1 Ligation of DNA strands into plasmid

Taq DNA polymerase has a terminal transferase activity that adds an adenylate nucleotide to the 3' end of the PCR product. The TOPO cloning vectors (Invitrogen) possess a 3' deoxythymidine overhanging nucleotide and topoisomerase which allows PCR products to be ligated efficiently into the vector.

In brief, using Invitrogen's TOPO® TA cloning kit (cat #46-0800) and following the manufacturer's instructions (25-0185-G-08-2701) 2 µl PCR reaction product, 1 µl salt solution, 2 µl ultra-pure water and 1µl TOPO cloning vector (pCRII) were mixed and incubated at room temperature for 20 minutes. This produced the pCRII:ECR transgenic construct which were then used to transform *E. coli* bacteria and the population was expanded.

2.7.2 Transformation and selection of colonies

Briefly, a vial of competent cells at -80°C was placed on ice for 5 minutes at which stage the cells were gently mixed. 2ul of the mixture from the previous ligation step, containing the ligated plasmid, was added to the vial of competent bacteria and the manufacturer's protocol was followed. Specifically, incubation on ice was allowed to progress for 30 minutes prior to heat-shock for 30 seconds at 42°C. Following the addition of 250ul of SOC medium at room temperature, incubation was allowed to proceed for 1 hour in a shaking incubator @ 37°C/ 250 bpm. Controls without plasmid were performed to test the efficacy of the antibiotic and to test for the presence of an antibiotic-resistant DNA strain.

Using a flamed, glass spreader 10µl and 100µl of cells were spread onto LB agar plates doped with ampicillin at 50 µg/ml (Table 22 Appendix). The inverted plates were incubated

overnight, in normal atmosphere, at 37°C to establish suitable colonies of bacteria-containing plasmid:ECR constructs.

2.7.3 Selecting and expanding colonies

Colonies were selected by STET prep but more typically by PCR-screening of a smear from an individual colony and using the original primers that had been used to amplify the DNA from the genomic DNA. The PCR protocol previously described was followed, except that the smear replaced the 1-10µl of DNA template. The resulting PCR output was checked for molecular weight by agarose gel electrophoresis.

2.8 STET prep (Sucrose, TRIS, EDTA, Triton) or Boiling Lysis Plasmid Extraction.

STET preps were also performed on occasion to establish the presence of plasmid. All steps performed at room temperature, unless stated otherwise.

1 ml of the overnight culture was pelleted by centrifugation for 1 minute at top speed and the ensuing supernatant was discarded.

The bacterial cell pellet was re-suspended in 350 µl STET buffer (8% w/v sucrose, 5% v/v Triton X-100, 50mM EDTA, pH 8.0, 50mM Tris-HCl, pH 8.0). 25ul 1 mg/ml lysozyme was added and mixed by vortexing. Following five minutes of incubation at room temperature, the bacteria were lysed at 95°C for 3 minutes and pelleted by centrifuging at top speed for 10 minutes leaving the plasmid in the supernatant. The debris pellet was dragged out of the tube with a sterile yellow pipette-tip and discarded. 350 µl isopropanol (2-propanol) was added to the supernatant and mixed by inversion to precipitate the plasmid DNA, which was then pelleted by centrifuging for 7 minutes at top speed. The supernatant was then decanted and the remaining isopropanol was allowed to drain and/or evaporate at room temperature for 10 minutes by inverting tubes on tissue in a rack. The plasmid was re-

suspended in 50 µl sterile DIW and heated at 55-60 °C for 10 minutes. Agarose gel electrophoresis was run to identify the presence of the plasmid.

2.9 Agarose gel electrophoresis

Throughout the work performed in this thesis, agarose gel electrophoresis was used to help determine the purity and estimate the concentration of the DNA, but mainly to extract the required DNA fragments for further processing.

2.9.1 Making and running the agarose gel.

Agarose gels were made by heating to boiling, agarose powder (Sigma) in 1 X TAE (Table 22 Appendix) and casting the gel with 5 mm wells, as per standard protocols (Sambrook et al., 1989, Cold Spring Harbor Protocols, 2006).When cooled the gel was submerged in fresh 1 X TAE in the electrophoresis gel tank. Gels of different agarose concentrations were made depending on the DNA fragment sizes expected. A concentration of 0.8% agarose was used to measure high molecular weights such as plasmids (thousands of bp) and up to 1.8% agarose was used to measure fragments of low molecular weight (tens to hundreds of bp).

Sufficient sample to provide between 10ng and 2 µg of dsDNA was mixed with appropriate amounts of loading buffer and made up with water. The loading buffers used were either BlueJuice 10X Cat. No.10816-015 for general use; TrackIt™ Cyan/Orange Loading Buffer 6X Cat. No.10482028, for DNA lengths 10bp - 1kbs an, TrackIt™ Cyan/Yellow Loading Buffer 6X Cat No. 10482035 for DNA lengths of 100bp – 10kbp, all from Invitrogen) and SYBR® Green I nucleic acid gel stain (SigmaAldrich Catalogue number S9430) at a final concentration of 2X. 20ul was loaded into the pre-formed wells in the agarose gel.

A DNA molecular weight marker (DNA ladder 1kb plus, Cat. No. 10787-018, 100 bp molecular weight marker Cat. No. 15628-019. TrackIt™ 50 bp DNA Ladder Cat. No. 10488-043, TrackIt™ 100 bp DNA Ladder Cat. No. 10488-058, all from Invitrogen) mixed with SYBR®

Green I nucleic acid gel stain was also loaded in separate lanes on the gel, if they were required.

A voltage of 5V/cm was applied across the gel and when the marker dye had migrated sufficiently the gel was viewed under a Dark Reader blue light transilluminator (excitation wavelength centred at ~400 to ~500 nm).

2.10 Recovery of DNA from Agarose Gels

2.10.1 Excision of DNA-containing gel

Whilst illuminated by the Dark Reader the fluorescing piece of gel containing the required DNA was excised from the agarose gel slab using a clean blade. The DNA was extracted from the agarose gel using either SV Gel clean-up kit where the manufacturers' protocol was followed by glass-fibre shattering. For glass-fibre shattering the DNA fragment was recovered by shattering the gel with centrifugation for 2 hours at 13,000 RPM through a punctured, fibreglass-filled 0.5 ml centrifuge tube mounted in a 1.5 ml centrifuge tube. The DNA fragment passes through to the 1.5 ml collecting tube whilst the agarose is retained by the fibreglass. This was followed by ethanol precipitation of the DNA fragment.

2.10.2 Ethanol Precipitation

Ethanol precipitation was performed by mixing one volume of DNA mixture, 2 volumes of ice-cold molecular-grade ethanol and 0.1 volumes of 3M Sodium Acetate (final concentration 0.3 M, pH 5.2) and agitating to ensure thorough mixing. This was stored for several hours at -80°C or overnight at -20°C or -80°C. Following the cold storage the mixture was centrifuged for 20 minutes at ≥ 25,000 x g, the supernatant then removed by pipetting and the pellet washed once in 70% molecular-grade ethanol diluted in ultra-pure H_2O and centrifuged at ≥ 25,000 x g for 5 minutes. The supernatant was removed as before and the

ethanol allowed to evaporate at room temperature. The pellet was the then re-suspended in a suitable volume of ultra-pure H_2O to result in the desired DNA concentration.

2.10.3 SV Gel clean-up kit protocol

Promega Wizard SV Gel & PCR clean-up system Cat. #A9281 was used to extract the DNA from the agarose gel. This kit contains a Membrane Binding Solution (MBS) which dissolves the gel and assists the DNA to adhere to the membrane in the Minicolumn Assembly. A Membrane Wash Solution is provided to remove impurities and nuclease-free water to elute the DNA from the membrane. Briefly, the excised gel slice was dissolved at a ratio of 1µl MBS to 1mg of gel at 50–65°C or for a PCR amplification an equal volume of PCR reaction and MBS. The mixture was centrifuged through a Minicolumn Assembly at 16,000 × g for 1 minute to transfer DNA to the membrane. The membrane was washed twice by centrifuging Membrane Wash Solution at 16,000 × g. Following evaporation of residual ethanol the DNA was eluted from the membrane in 50µl of nuclease-free water.

2.10.4 Phase-lock gel Phenol/Chloroform Extraction

This process removes proteins from the nucleic acid. The nucleic acids remain in the aqueous phase and proteins separate into the organic phase or lie at the phase interface. Phase Lock Gel acts as a barrier between the organic and aqueous phases, allowing the nucleic-acid-containing phase to be more easily decanted or pipetted off.

PCR product was mixed with an equal volume of 25:24:1 (v/v) phenol, chloroform, isoamyl-alcohol, added to PLG tube and centrifuged for 5 minutes at 13,000 rpm. The nucleic-acid-containing aqueous upper phase was then decanted and ethanol precipitated.

2.11 DNA Measurements

DNA concentration measurements were performed either using the standard protocol of absorbance at 260nm using a spectrophotometer or by using Invitrogen's Qubit Quantitation Platform (Cat. No. Q32857).

2.11.1 Qubit Quantitation Platform

Invitrogen's Qubit Quantitation Platform together with its associated assay kits was used to quantify amount of DNA. Two Quant-iT™ kits are available. The Quant-iT™ dsDNA BR Assay Kit (Cat. No. Q32850) is used for for normal DNA concentrations ranging from 2ng to 1000 ng whilst the Quant-iT™ dsDNA HS Assay Kit (Cat. No. Q32851) is used for low concentrations ranging from 0.2ng to 100ng. The manufacturer's protocols were followed for the Qubit Platform measurements. Briefly, sufficient working solution was made up by diluting the fluorescent reagent to X1 in buffer and the system was calibrated using supplied standards diluted in the working solution. The sample, diluted if required to the kit's operational range, was added to the working solution. The reading produced for the sample was then calculated using the following equation:

Concentration of sample = $QF\ X\ (\frac{200X\ DF}{x})$

Where:

QF = reading from Qubit fluorometer in ng/ml for HS and µg/ml for BR.

DF = Dilution factor

x = microlitres of sample used in assay.

2.12 Assay concentration of DNA and protein contamination by spectrophotometry

2.12.1 Concentration

DNA samples diluted in distilled water to give a concentration of between 10 ng/µl and 50 ng/µl and dilution factor noted.

Spectrophotometer was calibrated at 260 nm using distilled water. The optical density of the sample at 260 nm (OD_{260}) was obtained and following formula used to calculate the DNA concentration in ng/µl.

OD_{260} X 50 X dilution factor = concentration of DNA(ng/µl)

where:

OD_{260} = optical density of sample at 260 nm.

2.12.2 Calculate DNA purity

The optical density of the DNA was measured as above at 260 nm (OD_{260}) and also at 280nm (OD_{280}) and the following formula was applied:

DNA purity ratio = $\frac{OD_{260}}{OD_{280}}$

where:

OD_{260} = optical density of sample at 260 nm.

OD_{280} = optical density of sample at 280 nm.

A ratio of ≥ 1.6 is required for successful transfection.

2.13 Estimate DNA Concentration

Estimates of DNA concentration were also performed by comparing the intensity of the fluorescence from the DNA sample against the intensity of the fluorescence from a known quantity of molecular weight marker of approximately the same molecular weight.

2.14 Single and Serial Double Digest General Protocol

Restriction enzymes cut nucleic acid sequences at specific, usually palindromic motifs known as restriction sites. In double-stranded DNA the cuts may be blunt or have overhanging "sticky" ends. All digests used in this process resulted in overhanging ends.

2.14.1 Single Digest

Up to 17 µl of plasmid DNA, 10U of Restriction Enzyme (e.g. NotI – Invitrogen), 2ul compatible buffer (e.g. REact 3) were mixed in a thin-walled PCR tube and incubated for 3 hours at 37°C.

The mixture was then run on a gel of suitable density (e.g. 0.8%) with either a MW marker and/or undigested plasmid in separate lanes. Because a linearized plasmid moves through an agarose at a different rate to a circular plasmid an apparent molecular weight size shift against the undigested plasmid indicates a successful cleavage by an enzyme.

2.15 Double Digest - Serial

2.15.1 First Digest

Same protocol as 'Single Digest' above was used.

2.15.2 Ethanol precipitation

The DNA was then ethanol precipitated to remove the first enzyme and its associated buffer. The DNA was then re-suspended in 17 µl H_2O and 1ul of second restriction enzyme (e.g.

SpeI) and 2ul of its compatible buffer (e.g. REact 4) were added, mixed and incubated for 3 hours at 37°C.

2.15.3 Second Digest

The DNA was separated through electrophoresis on an agarose gel. The band of the correct size was excised and the DNA fragment was recovered and purified as previously described.

Note: When the composition of the two buffers allowed it, the ethanol precipitation step was eliminated by modifying the buffer used by the first restriction enzyme to meet the second restriction enzyme's requirements.

2.15.4 Double Digest – Simultaneous, using Fermentas Fast Digest Protocol

The Fermentas FastDigest® product allows for the simultaneous digestion of DNA by multiple restriction enzymes in a single buffer.

Briefly, 2µl purified plasmid (~ 0.2 µg plasmid), 1 µl of each FastDigest® restriction enzyme (e.g. NotI and SpeI), 14 µl ultra-pure water, 2µl 10X FastDigest® buffer were mixed and incubated at 37°C for 1 hour, followed by agarose-gel electrophoresis and extraction from gel.

2.15.5 Concentration of PCR product

The concentration of the PCR- amplified DNA fragments was increased by ethanol precipitation.

2.16 Cloning Strategy

Two cloning strategies were used, one allowed for directional cloning and the other was non-directional cloning.

For directional cloning the restriction sites plus extra base pairs were incorporated into the outer extremities of the primers being synthesised. The extra base pairs were required as restriction enzymes will not reliably cut if the recognition site is at the end of the DNA. The number of extra base pairs required is restriction-enzymes dependent.

DNA was excised by digesting at the restriction sites NotI and SpeI using appropriate restriction enzymes.

For non-directional cloning the putative enhancer was simply TA-cloned into pCRII-TOPO and following the overnight population expansion was excised from pCRII-TOPO using the NotI and SpeI restriction enzyme sites flanking the insertion site.

2.16.1 Cloning putative ECR into reporter vector

The required ECR sequence was amplified by either PCR or PCR and TA cloning as previously described in " PCR Amplification of DNA". The putative enhancer was then ligated into the p1229 reporter.

2.16.2 *Ligation of putative enhancer into p1229 vector.*

The putative enhancer was ligated into the p1229 reporter vector using Invitrogen's T4 Ligase (Cat. No. 15224-041) and the manufacturer's protocol for general cloning was followed (Invitrogen DOC. REV. 051502 p3). The Insert:Vector molar ratio was calculated with a target of 3:1 to be used in the protocol. For cohesive ends 3-30 fmols of Vector ends and 9-90 fmol Insert ends are required. In brief, for a 20µl reaction, reaction buffer (final concentration 1X) vector and insert DNA to a total of 0.01-0.1 µg (cohesive ends) were mixed with 1 unit of T4 DNA ligase and incubated for 5 minutes. This was then used to transform competent bacteria as described previously in "Transformation and selection of colonies".

2.17 Shrimp Alkaline Phosphatase (SAP) protocol

Shrimp Alkaline Phosphatase (SAP) (Fermentas Catalogue Number EF0511) catalyses the release of 5'-phosphate groups from DNA and reduces the risk of the vector re-annealing. Although typically only required for matching or blunt ends, due to problems encountered, this protocol was used on a number of occasions.

2.17.1 SAP treatment Protocol for dephosphorylation of DNA 5'-termini

17 µl of the solution containing the DNA sample, 10X reaction buffer 2 µl, and SAP (1 µl at 1 unit/1 µl) were incubated at 37°C for 30 minutes and reaction was stopped by heating at 65°C for 15 min

2.18 Plasmid Purification

2.18.1 Plasmid prep

To produce plasmids suitable for electroporation, with a 260nm:280nm λ ratio >1.6, overnight cultures of bacteria containing the plasmids were purified using plasmid prep kits from different manufacturers depending on availability. These were PureYield™ Plasmid prep (Promega), GenElute™ Plasmid (Sigma-Aldrich) and Qiagen's (Cat: 12143) products. The mini, midi and maxi versions of their products were used depending on the amount of culture being purified. The manufacturer's protocols were followed. The general methodology comprises lysing the bacteria to release the plasmid, recovery of the plasmid on a membrane, followed by elution to release the plasmids. Ethanol precipitation was the used to achieve ~5ug/µl suspended in PBS + 1mM Mg^{2+} required for transfection.

2.18.2 Sequencing

Promoter sequences, such as SP6, T3 or T7, are incorporated into many plasmids, bracketing the insertion site to allow sequencing from either end of the insert. Sequencing of the plasmids was performed by Lark Technologies (now Beckman Coulter) using their most simple offering i.e. Silver service on Perkins Elmer Biosystems DNA analysers from the appropriate promoter(s). Sp6 and T7 promoters were used in pCRII-TOPO and although p1229 contains the T3 and the T7 promoters, only T7 is sufficiently close to the insertion site to permit sequencing.

2.19 AB290 Rb anti-GFP

Same procedure as for MNR2 was followed except that the primary antibody, Ab290, was used at 1:2000 and that Alexa 555 rat anti-rabbit at 1:250 was used as the secondary antibody.

2.20 Injection and transfection by electroporation.

2.20.1 Age of embryo

The most suitable Hamburger-Hamilton (HH) stage to inject the neural tube was found to be at about HH14. Earlier than this and the NT had not formed a closed central canal of sufficient length, later than this and the central canal of the neural tube becomes too small.

2.20.2 Plasmid Mixes for Injection

The plasmid mix was made of a mixture of two plasmids at a maximum plasmid load of 6µg/µl.

The IRES-GFP constitutively active plasmid (final concentration of 2ug/µl) and the p1229 chromagenic reporter plasmid containing the sequence under test (final concentration of

3ug/µl) were suspended in PBS +1mM Mg^{2+}. A small amount of Fast Green (FG) was added to assist visualization during the injection process.

The co-injection of the GFP-expressing vector allowed confirmation that the transfection process had been successful and to indicated where the blue chromogenic stain would be located. Because the GFP is visualized using a fluorescent-type microscope it avoided the chromogenic processing that necessitated the sacrifice of the embryo by extraction from the egg and processing for staining.

2.21 Accessing the Embryos

Two methods have been used in the experiments used in the development of this thesis; Hammocking and Windowing.

In both cases the eggs had been marked (usually with the start-of-incubation date), and had been placed in the incubator with this label uppermost. Because the embryos float to the highest part of the egg this indicated where it as anticipated that the embryo would be located.

The egg was removed from the incubator and rocked gently to dislodge the embryo as some embryos have a tendency to adhere to membrane.

The eggshell was sterilized by spraying it with 70% ethanol.

2.21.1 Hammocking Eggs:

Hammocks had been prepared by placing a piece of non-PVC cling-film over 2.5" length of plastic pipe of 3" diameter and secured with an elastic band to form a cup or hammock about 1.5" deep.

The egg, kept in the same orientation in which it had been incubated, was then cracked on the bottom and its contents emptied into the hammock. The embryo was then ready for

injection. A Petrie dish was then placed on top to maintain humidity and prevent ingress of dirt.

Due to yolk breakage, infection and a concomitant high mortality rate the hammocking method was abandoned in favour of 'windowing'.

2.21.2 Windowing eggs:

This windowing method removed the need to empty the egg out of the shell with its attendant risk of damage to the yolk, which becomes very delicate following fertilization.

As this process consists of cutting a hole in the eggshell a sufficient amount of the albumin had to be removed to prevent it overflowing from the hole.

The pointed end of the egg was punctured with the egg punch and a 19G hypodermic needle mounted on a syringe was introduced through the hole and directed below the equator mark to avoid the yolk, and between 2.5 and 4 ml of albumen was aspirated. This amount depended on the size and freshness of the egg, typically the larger and fresher the eggs the greater the amount that had to be aspirated.

An oblong of 2.5 cm by 1.8 cm was marked on the top of the egg, centred on where it was anticipated the embryo should be, i.e. on the part of the egg that was highest during incubation.

Using a 38 cm diamond cutting wheel mounted on a flexible shaft attached to a Dremel-type rotary multi-tool the marked oblong was cut through the shell (Figure 24) avoiding puncturing of the membrane if possible.

Figure 24 Egg being windowed showing shell cut through exposing the membrane. To top of picture is the cutting disc mounted in the chuck of the Multitool.

The dust from the cutting was removed from the egg and the membrane was cut through on the long (2.5cm) sides of the oblong with a scalpel. A suitable length of Sellotape® Magic tape (2.5 cm wide) was placed over partially cut oblong or "window". This was pressed on to the cut out portion of the shell and onto the shell at one of the narrow cuts to provide a hinge. The tape was not stuck down on the long sides nor at the other short end. The membrane at the free, short side was cut and the piece of shell, thus hinged on the other short side, was raised.

A check was made that sufficient albumen had been removed so that leakage would not occur, bubbles that may have formed in the albumin were burst with the hypodermic needle, and any membrane that could dip into the albumin was removed. The window was then closed and the Sellotape® was lightly pressed down all around to form a seal. Originally ordinary clear Sellotape® was used but subsequently Magic Scotch tape was used and this reduced the reduced the risk of eggshell breakage when manipulating the tape, presumably due to the lower adhesion of the adhesive on the Magic tape.

The puncture hole was sealed with a piece of laboratory tape (Shamrock brand) and the egg was then ready for injection.

2.21.3 Plasmid injection

7% India Ink (Pelikan Fount India ink) in PBS was sometimes injected under the embryo to help visualise it, but this appeared reduce survival of the embryo and was avoided if possible.

The membrane covering the embryo was the removed using a forceps.

A micropipette of borosilicate capillary glass tubing, with filament, (Warner Instruments, G150TF-3) had previously been pulled (Sutter Instruments Model P-87 Flaming/Brown pipette puller) and attached to a flexible mouth tube. These were then loaded with the plasmid mix. The neural tube or brain, as appropriate, of the chick embryo was pierced with the micropipette tip (Figure 25) and the plasmids injected into the central canal of the neural tube or the lumen of the brain using a small amount of buccal pressure.

Gold plated electroporator electrodes of 3mm or 5mm length (Harvard Apparatus, Inc. BTX Model 514(3mm) 512(5 mm)) were placed each side of the embryo with a gap of 5 mm and connected to the electroporator. A drop or two of L15 was sometimes placed on the electrodes to improve electrical contact.

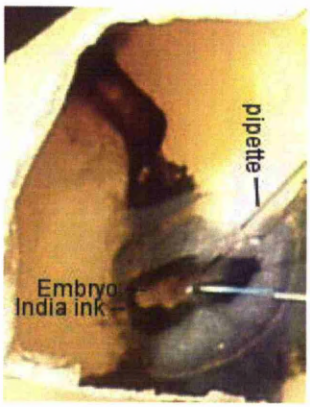


Figure 25 Neural tube of a chick being injected using a micropipette. Note this picture is of a windowed egg with the shell visible, and not an egg in a hammock.

2.21.4 Electroporation

Upon withdrawal of the micropipette the embryo was electroporated with a number of pulses (typically 5) at a determined voltage (ranging from 40V to 15V) with a pulse length ranging from (50msecs - 20msecs) duration, with a gap length ranging from random gap to 100 msecs.

The electric charge across the tissue causes pores in the cells to form and at the same time the plasmid DNA, which is negatively charged, is drawn towards the positively charged electrode and is drawn into the cells through the pores Figure 26.

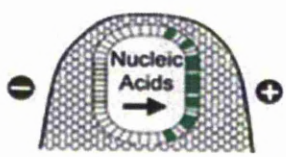


Figure 26 Diagram showing electroporation of neural tube. The plasmid having been injected into the lumen upon passing of current through the electrodes, is drawn towards the positive pole and transfects some of the cells (shown in green).

Because the DNA is drawn towards the anode the placement of the electrodes in relation to the tissue will alter the location of the tissue that will be transfected. When transfecting tissue of the neural tube with DNA injected into the central canal, when both electrodes are placed one on each side of the torso at the midline of the dorso-ventral axis the centre-point of electroporation will occur at the midline on the anode side. If the cathode is placed ventrally of the midline and the anode is placed dorsally the electroporation will occur in the dorsal half of the anode side and vice-versa if the anode is placed ventrally and the cathode dorsally (Figure 27). This manipulation can be used to advantage to target a particular location however it can unintentionally and lead to the undesirable result of only part of the lateral half of the neural tube being transfected.

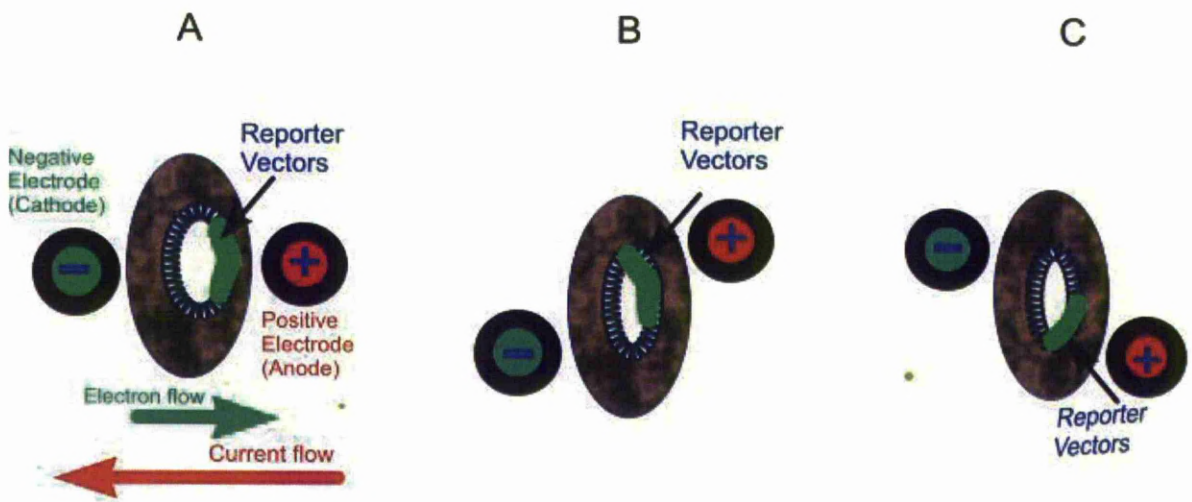


Figure 27 Targeting of region to be electroporated by positioning of electrodes, shown in a cross section of embryonic chick torso. (A) Both electrodes placed at midline results in the centre-point of electroporation occurring at the midline of the anode side. (B) Cathode placed ventrally and anode placed dorsally results in the centre of electroporated tissue being more dorsal. (C) Cathode placed dorsally and anode ventrally results in the centre of electroporated tissue being more ventral.

Two electroporators were used during the development of this thesis. The first electroporator was battery powered with limited controls (Digitimer Isolated Stimulator Model DS2). When using this model the voltage was typically set at 40V with a pulse length

of 50msecs. The trigger being a manual push-switch produced, by its nature, a random gap length between pulses. The only control of gap length was the very subjective, pressing the trigger button at a fast or slow rate.

The second electroporator was a more advanced, mains-powered, Ovodyne TSS20 with user-selected but machine-implemented gap timing. There is also a maximum current limit to help reduce tissue damage. Following the commissioning of this instrument the electroporation success increased.

The embryo was then covered, by a petri dish when the Hammocking method was used. When using the windowing system the shell fragment was taped down or if the shell fragment was lost, taping over the window. The embryo was then placed in a humidified incubator.

Two types of humidifier were used; the earlier experiments used a humidifier set at 37°C and passive humidification. The later experiments had the advantage of a specialised egg incubator system (Brinsea Ova Easy 380 with H22 Digital Humidity Management Module) set at 37.8°C and with an active humidity system set at 35%.

Following the incubation period, the embryos that expressed GFP fluorescence in the appropriate location, were harvested, fixed, permeabilized and stained for ß-galactosidase (*lacZ*) activity using X-gal as the chromogenic substrate as described below.

2.21.5 Harvesting Embryos

The embryos were extracted from the shell by severing the vitelline membrane and blood vessels from around the embryo by using sharp forceps. The embryo was then scooped up using curved-shank tweezers, a spatula or a small laboratory spoon. The embryos were then irrigated with Hank's or other buffered solution. The embryos were then cleaned up, using sharp forceps, by removing any remaining vitelline membrane and unwanted tissue. The

membrane surrounding the embryo and viscera were typically removed at this stage from the more advanced embryos. This clean-up was performed on the more delicate embryos that were at an earlier stage of development after they had spent some time in the fixative.

2.21.6 Fixing Embryos

The embryos were immersed in 4% paraformaldehyde fixative (Table 22 Appendix) for a period sufficiently long to maintain tissue integrity but not sufficient to destroy the ß-galactosidase enzyme activity. This period was typically from 1 to 4 hours, depending on the age of and size of the embryo. Following this treatment the embryos were ready for further processing such as staining and sectioning.

2.22 Whole-chick embryo staining protocol

Adapted from (Nagy et al., 2007)

2.22.1 Washing:

The paraformaldehyde was aspirated from the suitably fixed embryos. These were then washed in embryonic detergent rinse (Table 22 Appendix) three times for 15mins per wash on a rocking table (30 bpm) prior to staining.

2.22.2 Staining:

The embryos were submerged in staining solution (Table 22 Appendix), typically ~500 µl to 2 ml per embryo, depending on the embryos size.

2.22.3 Developing the stain:

Depending on the circumstances a number of protocols were used for developing the indigo stain. The embryos were either placed at 37°C with the stain expected to be visible within 5 hours; or overnight at RT on a rocking table (30 bpm), if no staining was visible further incubation at 37°C was undertaken. Some embryos were also stored for several days at 4°C, prior to incubation at 37°C.

When stain was visible the reaction was typically stopped in 70% Ethanol.

2.22.4 Mounting chick embryos on cork discs.

To cryo-protect the tissue/embryo it was immersed, until sinking, in the succession of 6%, 12% and 18% sucrose in working-strength Millonig's Phosphate Buffer (0.125M, pH7.4) with 0.1% NaN_3.

The required section of a cryo-protected embryo was placed in the desired orientation, in a dome of optimal cutting temperature (OCT)-type embedding compound (Bright Cryo-M-Bed medium) on a cork disc.

Iso-pentane was placed in a plastic beaker and suspended over liquid nitrogen in a Dewar's flask. The cork disc was then placed in the ultra-cold iso-pentane for about 1 minute to allow the OCT and tissue to freeze at which point it was immersed in the liquid nitrogen.

The tissue typically then stored at -80°C

2.23 MNR2 (Hb9) Antibody-staining of chick embryos.

Sections (10 micron) of chick neural tube (E5,E6,E7) were mounted on gelatine subbed slides,paraformaldehyde was quenched (15 minutes in 0.2M Glycine pH 7.4), non-specific antibody sites blocked in diluent (15-30 minutes) (Table 22 Appendix), stained with primary antibody at 1:5 in diluent (30mins – 1 hour at RT), dip washed (3 times) in HBSS, stained with Alexa 488 rabbit anti-mouse secondary antibody at 1:200 in diluent (1 hour at RT), dip washed (3 times) in HBSS and once in de-ionized water and cover-slipped using DAKO medium as adhesive.

2.24 DAPI staining

Nuclear staining was achieved by co-application of DAPI (1:100 in diluent) with the secondary antibody.

Chapter 3

3 Development of a Positive Control to demonstrate p1229-based system would function in chick.

The expression of genes by cells is under very stringent control. Of particular interest in this work are the *cis*-acting enhancers and/or repressors of gene expression. These regions bind the transcription factors that are produced by the cell and either increase or decrease the rate of transcription initiation.

One way to identify these regulatory regions is by is to identify regions of DNA that have been restrained over evolutionary time. These evolutionary conserved regions (ECRs) can then be tested as putative enhancers by insertion into a reporter vector.

3.1 Selection of positive control genes and enhancer.

Prior to using the chick-based model, as described in Chapter 1, to assay the ECRs for *LSAMP*-enhancer activity, a positive control, with known enhancer function, was required to show that we could make the proposed system and that it would be shown to function.

A number of criteria were specified for the positive control enhancer; it should be for a known promoter of a gene that is expressed early in development in the neural tube of chick and that it should reliably produce a strong, easily detectable signal.

The reason for selecting a gene expressed early in development was that *LSAMP* expression was to be investigated at this stage of development and the gene being used for the positive control should match this temporal expression pattern. The neural-tube location was desirable because it was in this location that the investigation into *LSAMP* was to commence.

A number of candidate genes and their known enhancers were investigated and of these genes *Pax3* and *Hb9* appeared to be the most promising. *Pax3* is a member of the paired box

(PAX) family of transcription factors and minimal enhancer elements sufficient for *Pax3* expression in neural crest cells had been identified (Milewski et al., 2004). Hb9 is a transcription factor with a 5', 125 bp region (Figure 28) demonstrated in E10.5- E11.5 mouse (Figure 29), to be both sufficient and necessary to drive reporter gene expression in the appropriate location, i.e. the spinal motor neurons. Reporter vector, p1229 was also selected for testing.

3.2 The *PAX3* gene was not required for proof-of-concept testing.

Because results were obtained using the *Hb9* enhancer in p1229 no further reference to Pax3 or PGL3 will be made except to say that the equivalent chicken *PAX3* enhancer sequence was identified using cross-species homology as described in detail for *Hb9*, but the *PAX3* enhancer region proved difficult to amplify from whole chick genomic DNA. Although this was eventually achieved using nested PCR and the reporter plasmid subsequently cloned, the positive results obtained from the *Hb9* ECR obviated the need to pursue the *Pax3* enhancer and work on it and other genes was abandoned (data not shown).

3.3 The *Hb9* gene was used for proof-of-concept testing.

The *Hb9* gene encodes a nuclear protein, which contains a homeobox domain and is a transcription factor. Mutations in this gene can lead to sacral agenesis (Li et al., 1999, Ross et al., 1998). *Hb9 i*s expressed by spinal motor-neurons in the developing central nervous system and is a marker for motor neurons (Tanabe et al., 1998, Arber et al., 1999). Previous studies had shown that in mouse a 9kb 5' promoter region was sufficient to drive gene expression in motor neurons *in vivo* (Arber et al., 1999, Wichterle et al., 2002)) and this was further refined to a 3.6 kbp segment of this 9kbps by human to mouse homology (Nakano et al., 2005)). Transgenic mice generated with this 3.6 kbp segment inserted into the reporter vector BGZA, which contains a LacZ gene under a human β-globin minimal promoter, were

shown to express β-Gal in the same regions of the ventral spinal column as for the 9kbp segment. Comparing mouse, human and pufferfish (*Fugu rubres*) revealed a 438bp region consisting of a non-contiguous 313bp and 125bp fragment. Zebrafish (*Danio rerio*) were then found to have two identical copies of the 125bp region, but no 313bp region. Transgenic analysis showed the 125bp fragment was sufficient to drive *LacZ* expression in the spinal motor neurons and therefore an enhancer (Nakano et al., 2005). The 313bp region demonstrated no such function, indicating that the 313bp had no enhancer function at this stage. Furthermore disruption of a Hox/Pbx1 consensus binding sequence (TGATnnAT, where n represents any nucleotide) within the 125bp sequence destroyed the enhancer effect (Nakano et al., 2005).

The reverse complement of TGATnnAT (ATTCATCA, where TC is in the location specified by nn) occurs within the consensus sequence for the enhancer HNF1_C (ttgtCAATTCATCAGCt, where upper case letters represent the core sequence within the enhancer's consensus sequence and nn is any nucleotide).

(A)
agagtggttagctgatgaattgacaaaaactaatcagctttattgggaaacaggtttaaggg
cacggacgtgtcaataacgctcagcctgacccctcttccattagctxaggcaggctgatta
ga

(B)
tctaatcagcctgcctgagctaatggaagaggggtcaggctgagcgttattgacacgtccg
tgcccttaaacctgtttcccaataaagctgattagttttgtcaattcatcagctaaccact
ct

(C)
tgaataaatttaagcaggctaattaatatataaactagctcaatttgtcaagttgatttgta
ttttagttaattgtgaaagtaattaccacatggtcaaattaacagctttctggaaatgacca
agcctgaggttttatttccttcctgggtgaagaaaattcatttttccaagctcttgatgtga
tgaataaaagtcataaatctgggtgattggtgcaggcagagtctaaatggcttcatatttca
ttttaggtttaatagaaatattcatgctctgttttaatgaaattaaattgaaggggatggg
gct

Figure 28 Sequences investigated for enhancer function taken from Figure E (Nakano et al., 2005). (A) 125 bp sequence shown to act as an enhancer. Highlighted in yellow is the sequence highly related to the Hox/Pbx consensus binding sequence (TGATnnAT, where 'n' is any nucleotide), disruption of which was found to destroy the enhancer effect of the 125bp sequence. (B) the reverse complement of the 125bp sequence shown in A. (C) The 313bp mouse sequence not shown to act as an enhancer at this stage of development. Note the repeat of one unit of the Hox/Pbx consensus binding sequence as indicated in (A) above in this 313bp structure.

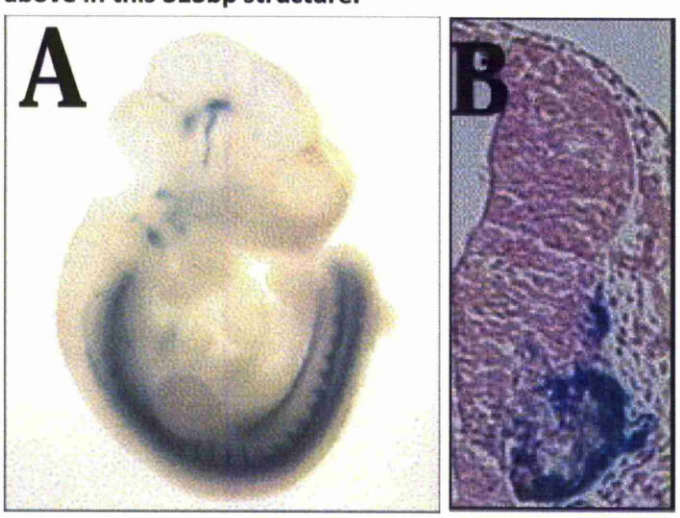


Figure 29 The expression of β-galactosidase is detected by blue staining in the spinal motor neurons of a transgenic mouse staged at (A) E10.5 whole mount and (B) 11.5 section of neural tube. Image taken from Fig 5 (Nakano et al., 2005).

3.4 Aim of chapter

The aim of this part of the project was to develop a positive control by demonstrating that reporter vector p1229 could be used successfully to detect if a DNA fragment was an enhancer of gene expression. This was to be done by identifying a DNA sequence that had been proven to be an enhancer, preferably in chick, but failing that, in another species. If the sequence identified was not in chick then the equivalent sequence in chick was to be located by a bioinformatics method such as BLAST (Altschul et al., 1990, Johnson et al., 2008). The enhancer sequence was to be cloned into a reporter vector and transfected into embryonic chick. The technique would be validated if the reporter gene expression could be detected in the expected location in the chick.

An additional aim of the experiments described in the chapter was to develop a methodology that would allow the chicks to be injected, electroporated and survive for the sufficient time to allow reporter gene expression.

RESULTS

3.4.1 Synthesis of the ECR1-Hb9:p1229 construct.

3.4.1.1 Identification of the mouse homologue enhancer sequence in chick.

The location of the Hb9 125 bp enhancer sequence (Nakano et al., 2005) as shown in Figure 28 was located by BLAST using Mouse genome resources (Altschul et al., 1990) showed the location of the 125 bp sequence to be ~7.5kbp bp upstream of the start of Hb9 (Figure 30). The ECR Browser (Ovcharenko et al., 2004) was used to compare this region using mouse as the base organism and chick and human as the target organisms (stringency; 100 bp length

70% similarity) to show that the ECR Browser could be used to indicate the location of putative enhancers. The ECRs so identified were 344bps in length at 77.6% identity in mouse with the chick ECR being slightly shorter at 339 bp (Figure 31).The mouse/chick/human results are shown graphically (Figure 32) and the region upstream from the Hb9 TSS with the ECR containing the enhancer is identified. Note that there are 8 ECRs upstream of ECR1 in human that do not appear in chick (marked * Figure 32). This is a good example of how using an evolutionary more distantly organism can highlight important and more highly constrained sequences. In this case it serves to highlight the ECR of interest, i.e. the ECR containing the 125 bp sequence, but in other circumstances important data could be lost as the ECRs highlighted in human may also be important. The downstream ECRs, being constrained, are also probably important at a fundamental level. Although these other ECRs may have a function in the control of Hb9 or some other gene, as they do not contain the 125 bp enhancer they are of no interest in this search for a positive control.

The mouse enhancer sequence was further compared against other species (Figure 33). It can be seen how the region, centred at the 344bp ECR in chick, with mouse as the base species, has, unsurprisingly, the greatest similarity to rat. In human the ECR appears as a discrete element, unlike rat, and has a slightly higher similarity with mouse. Frog and chick appear very similar to each other. Of the organisms compared here, puffer-fish exhibits the least homology.

```
Features flanking this part of subject sequence:
   7793 bp at 5' side: motor neuron and pancreas homeobox protein 1
   83323 bp at 3' side: ubiquitin-protein ligase E3C

 Score =   231 bits (125),  Expect = 8e-59
 Identities = 125/125 (100%), Gaps = 0/125 (0%)
 Strand=Plus/Minus

Query  1         AGAGTGGTTAGCTGATGAATTGACAAAAACTAATCAGCTTTATTGGGAAACAGGTTTAAG  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  14028526  AGAGTGGTTAGCTGATGAATTGACAAAAACTAATCAGCTTTATTGGGAAACAGGTTTAAG  14028467

Query  61        GGCACGGACGTGTCAATAACGCTCAGCCTGACCCCCTCTTCCATTAGCTAGGCAGGCTGA  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  14028466  GGCACGGACGTGTCAATAACGCTCAGCCTGACCCCCTCTTCCATTAGCTAGGCAGGCTGA  14028407

Query  121       TTAGA  125
                 |||||
Sbjct  14028406  TTAGA  14028402
```

Figure 30 BLAST of 125bp region against mouse. The 125 bp sequence is identified as 'query and the mouse as 'subject'.

A)

ECR :: Evolutionary Conserved Region	**Mouse**$_{[mm8]}$ - **Chicken**$_{[galGal3]}$ **ECR**
Type	**344** bp, **77.6%** identity
Location	**chr5:29816686-29817029**$_{[mm8]}$

B)

Alignment

```
                10        20        30        40        50
Mouse     AAaaCatGTatTACACtGCGACATTCTGTAaATGAGATAATGATACATAAACCCga-TGA
          ||   |   ||   ||||| |||||||||||||| |||||||||||||||||||||||    |||
Chicken   AAcgCgcGTccTACACcGCGACATTCTGTAcATGAGATAATGATACATAAACCCcccTGA
               4590      4600      4610      4620      4630      4640

         60                  80        90       100       110
Mouse     TcGATGTGACgggCCtGGatGtcttCTCTAAATtAGCtGCtCaCATCGACTGCTaATA-T
          | ||||||||    || ||  |    |||||||| ||| || | |||||||||||| ||| |
Chicken   TaGATGTGACaccCCcGG--GatggCTCTAAATgAGCcGCcCgCATCGACTGCTgATAaT
               4650      4660      4670      4680      4690

         120       130       140       150       160       170
Mouse     tGtGaGTTTATGAAAGCAATTTCGGTGCaAtCgGCGagAGGGTCTtCTtacTcTAATCAG
           | | |||||||||||||||||||||| | | |||  ||||||| ||    | |||||||
Chicken   gGcGcGTTTATGAAAGCAATTTCGGTGCgAgCaGCGc-AGGGTCTgCTcggTgTAATCAG
       4700      4710      4720      4730      4740      4750

         180       190       200       210       220       230
Mouse     cCTGCctaGCTAATGGAaGaGGgggtCaGGCtGAGcGtTATTGACACgtCcgTGCCCTTA
           ||||   ||||||||| | ||    | ||| ||| | |||||||||   |  ||||||||
Chicken   gCTGCtcgGCTAATGGAcGgGG----CgGGCgGAGaGcTATTGACACtcCgcTGCCCTTA
        4760      4770      4780           4790      4800      4810

                  250       260       270       280       290
Mouse     AaCCTGTTTCCCAATAAAGCTGATTaGTTTTTGTCAATTCATCAGCTaACCaCtCTgGCT
          | |||||||||||||||||||||| |||||||||||||||||||||| ||| | || |||
Chicken   A-CCTGTTTCCCAATAAAGCTGATTgGTTTTTGTCAATTCATCAGCTcACCgCcCTcGCT
             4820      4830      4840      4850      4860      4870

         300       310       320       330       340
Mouse     GgaAacaaATCAAAGCtGtATTTGtTCAataGAAg-CAACAAGTCaC
          |  |    |||||||| | ||||| |||    |||  ||||||||| |
Chicken   GcgActgcATCAAAGCcGgATTTGcTCAgggGAAatCAACAAGTCgC
             4880      4890      4900      4910      4920
```

Figure 31 Details from the ECR Browser grab function (base organism, mouse; target organism, chick) showing A) the length of the mouse ECR and its percentage similarity to the chick sequence and the genomic location of the mouse sequence (mouse, because it was the base sequence). B) Indicates the alignment of the ECRs between mouse and chick with the murine 125 bp region highlighted in turquoise.

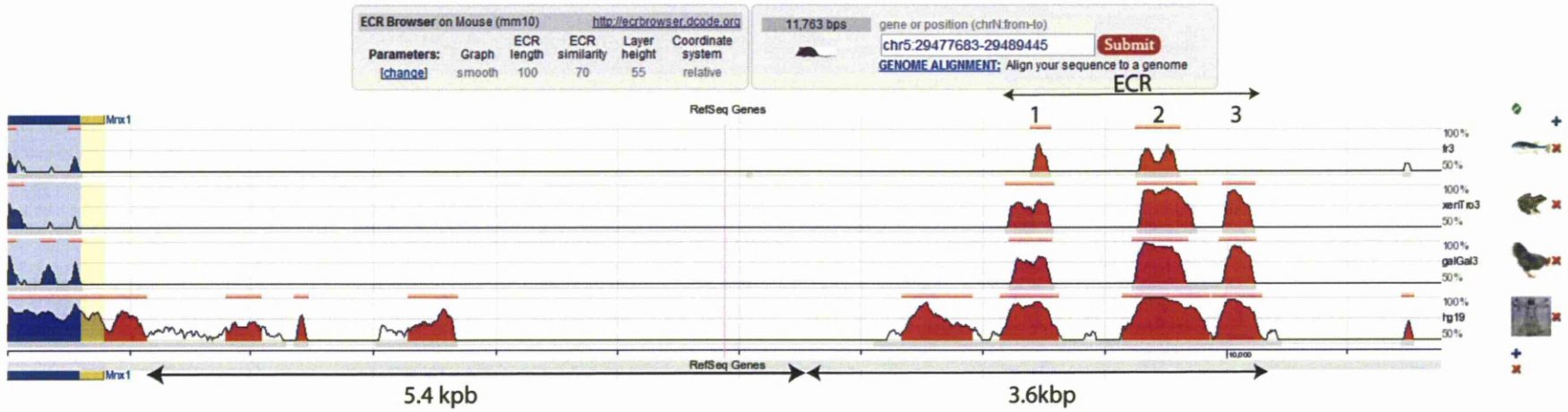


Figure 32 ECRs upstream of Hb9 (Mnx1)

Mouse as base organism and pufferfish (top line) ,frog, chicken and human (bottom line) as target organisms. Hb9 gene to left – marked as Hlxb9. ECR1 contains the ~125bp enhancer. ECR2 contains the 313bp region shown not to act as an enhancer. ECR3 was not investigated by Nakano – it appears after pufferfish in evolution. Note in human the 8 ECRs (marked *) upstream of ECR1 that do not appear in chick and the 2 ECRs downstream that appear in both chick and human. Location of the 3.6kbp and 5.4 kbp regions are indicated.

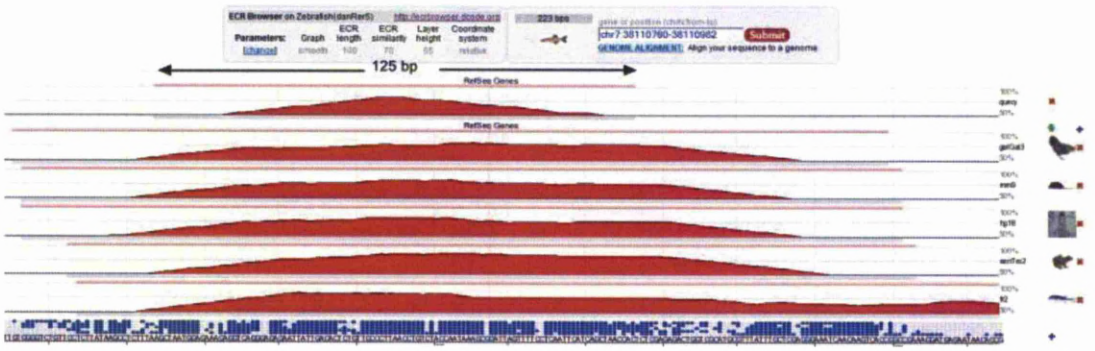


Figure 33 ECR Browser showing 125bp sequence identified as driving Hb9 expression in motor neurons .

125bp aligned against zebrafish (top line). Cross-species homology of ECR showing chicken, mouse human, frog and pufferfish. Stringency at default (100bp/70%). 100bp of the 125bp are constrained in puffer fish (Fugu).

3.4.2 Designing of primers to amplify ECR1 from whole-chick DNA.

Suitable primers sequences were generated using Primer3 (Rozen and Skaletsky, 2000) (release 1.0b) using default settings with the 120bp sequence from chick (Figure 31) marked as source sequence. To provide the optimal primers for the sequence, Primer3 selected a 215 bp region (Figure 34 and Figure 35).

	start	len	tm	gc%	any	3'	seq
1 LEFT PRIMER	9	20	58.00	35.00	6.00	3.00	CGCGTTTATGAAAGCAATTT
RIGHT PRIMER	223	20	59.67	50.00	3.00	2.00	CTTGTTGATTTCCCCTGAGC

PRODUCT SIZE: 215, PAIR ANY COMPL: 6.00, PAIR 3' COMPL: 0.00

Figure 34 Primer3 results chick Hb9 ECR showing the forward and reverse primers.

CGCGTTTATGAAAGCAATTTCGGTGCGAGCAGCGCAGGGTCTGCTCGGTGTAATCAGGCTGCTCGGCTA
ATGGACGGGGCGGGCGGAGAGCTATTGACACTCCGCTGCCCTTAACCTGTTTCCCAATAAAGCTGATTG
GTTTTTGTCAATTCATCAGCTCACCGCCCTCGCTGCGACTGCATCAAAGCCGGATTTGCTCAGGGGAAA
TCAACAAG

Figure 35 The 215 bp chick sequence to be generated by PCR showing the Forward primer (green highlight), Reverse primer (red highlight) as identified by Primer3 and the 120 bp putative Hb9 enhancer (turquoise highlight)

3.4.3 Amplification of Hb9's ECR1 DNA sequence.

The primers were synthesized, however amplification by PCR from the whole genomic DNA proved difficult and many attempts at a succession of annealing temperatures and denaturing, annealing and elongation times were tried, with varying salt concentrations and with and without adjunct. The problems in obtaining PCR product were probably due to attempting to amplify a very small fragment from genomic DNA and the stochastic nature of the initial annealing stages. The 215 bp region was eventually amplified by gradient PCR, as specified in Materials and Methods, from whole chick genomic DNA with the T_A temperature range set from 55°C to 65°C (Table 10). The PCR reaction used was as specified in Materials and Methods for 25 µl PCR reaction except that 1.25 µl H_2O was

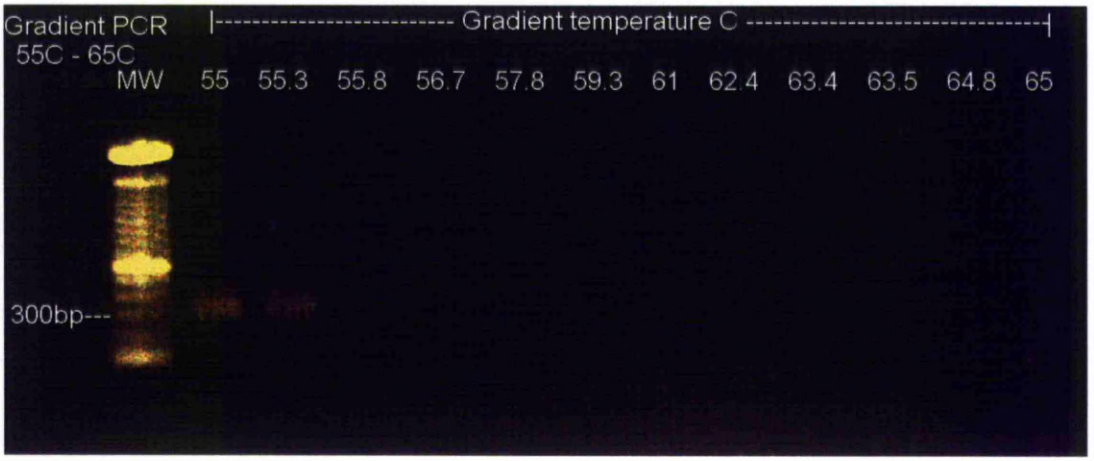
replaced by DMSO (5% v/v) as an adjunct and 63 ng whole-chick gDNA was used as the template .

The PCR reaction product was visualized by agarose gel electrophoresis (Figure 36 A) and track 1 (55°C) and track 2 (55.3°C), having produced bands at about the expected size were excised, purified, cloned into the pCRII vector, bacteria were transformed and expanded. Colonies were subsequently screened by PCR (T_A @ 55°C) using the original primers and the results visualised by agarose gel electrophoresis (Figure 36 B). The pCRII plasmid from four colonies (1,2,A and B) was extracted and purified by plasmid preparation. Sequencing from the Sp6 and T7 promoters showed the Hb9 enhancer had been cloned into pCRII, in colonies A,B and 1 the Hb9 enhancer was inserted in direct orientation but was reversed in colony 2.

Cell	1	2	3	4	5	6	7	8	9	10	11	12
t/°C	55	55.3	55.8	56.7	57.8	59.3	61	62.4	63.4	63.5	64.8	65

Table 11 T_A temperature gradient range of 55°C to 65°C of PCR used for Hb9 amplification from whole-chick gDNA.

(A)



(B)

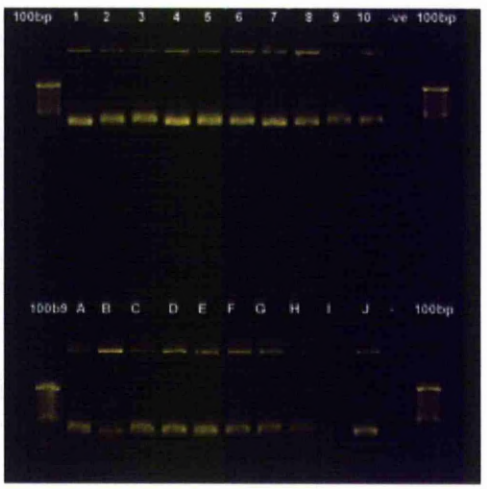


Figure 36 (A) 1.3% agarose gel showing the two bands of approximately the expected 215 bp that were obtained at 55°C and 55.3°C, by the Hb9 primers previously described, using Gradient PCR (temperature range 55°C to 65°C) on whole chick genomic DNA as a template. The MW marker is 100bp (Invitrogen) but probably due to the addition of SYBR Green to the DNA it ran a little slower than the naked DNA would. The track at 56.7°C was to be negative control, there being no gDNA in its mix, however as no band was produced at 55.8°C this functionality was negated.

(B) 1.3% agarose gel of the output from 10 colonies from the 100ul-spread plate (1-10) and the 10ul-spread plate (A – J) of pCRII plasmids, screened by PCR for inclusion of Hb9 ECR. All colonies, except colony I, produced a band of approximately the expected size.

The Hb9 enhancer was serially digested from pCRII for insertion into the p1229 chromogenic vector. The first digest was with restriction enzyme NotI, then following ethanol precipitation, pelleting and re-suspension in water the linearized plasmid was digested using SpeI. The reaction was run on a 1.3% agarose gel and the resulting 200bp band, visualised with SYBRGreen (Figure 37 A), was excised and the DNA was extracted. The DNA was ethanol precipitated and assayed as being 6.6 ng/µl.

The p1229 vector (2µl @ 342ng/µl) was double digested and visualised on 0.8% agarose gel. The size shift between digested (Figure 37 C) and undigested p1229 (Figure 37 D) is evident. This was excised, and processed to clean it up.

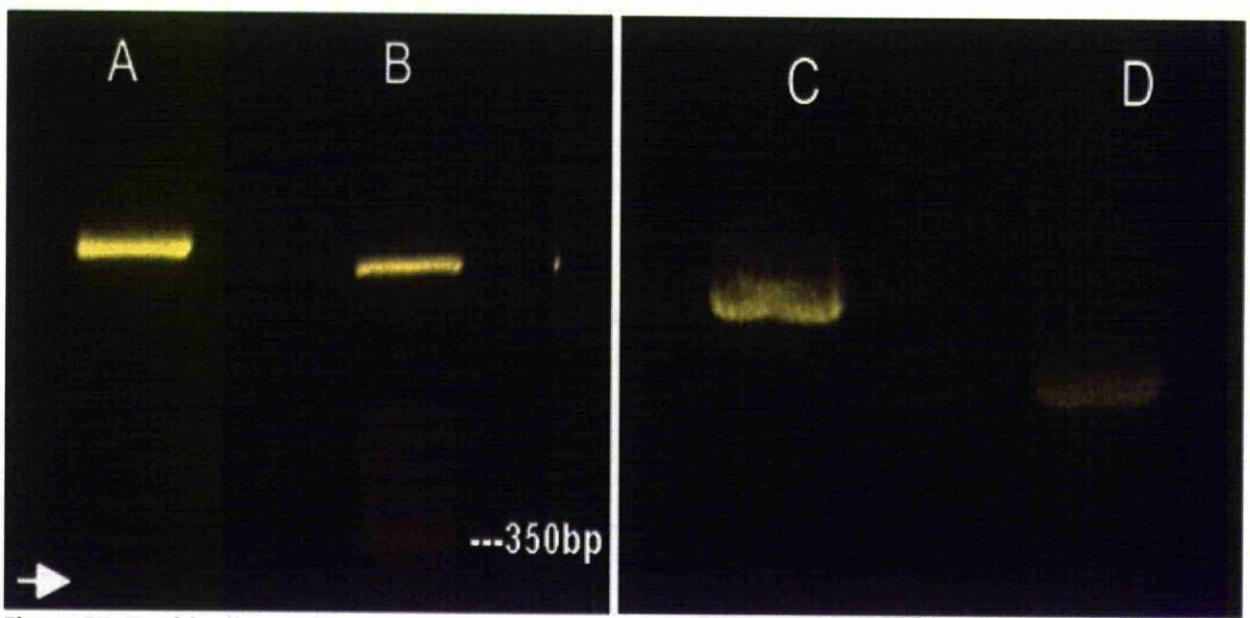


Figure 37 Double digests

(A) The 200bp Hb9-ECR (very faint, arrowed) double digest from pCRII with NotI and SpeI restriction enzymes.

(B) Molecular weight marker with 350 bp band identified.

(C) Plasmid p1229 double digested by NotI and SpeI with the size shift evident when compared with

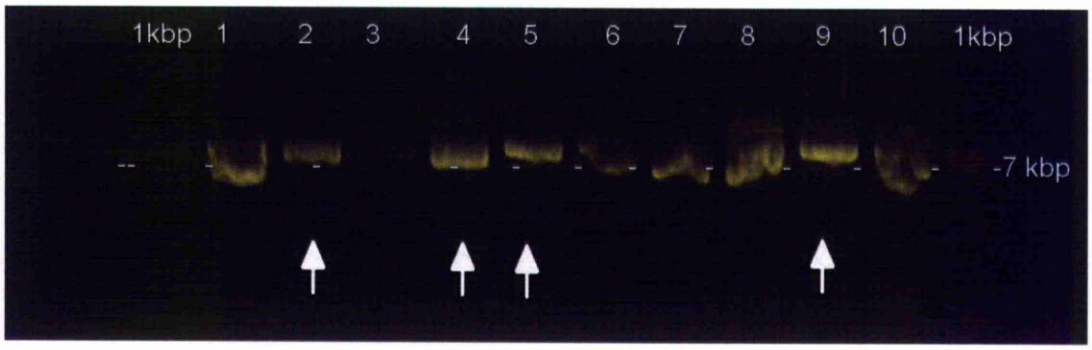
(D)The undigested form.

3.4.4 *Ligation of the Hb9 ECR1 sequence into reporter vector p1229 to generate the ECR1-Hb9:p1229 construct.*

Ligation of the Hb9 enhancer into the p1229 vector was performed. Following initial STET PREP verification of 10 colonies for the presence of plasmid (Figure 38 A) and subsequent double digest with NotI and SpeI) of 4 colonies (Figure 38 B), 2 colonies were purified by plasmid prep and sequenced (Figure 39). The sequencing indicated perfect homology between the desired and actual sequence.

These two colonies were expanded in overnight culture and the p1229 plasmid was extracted and purified and concentrated to 10.6 mg/ml.

(A)



(B)

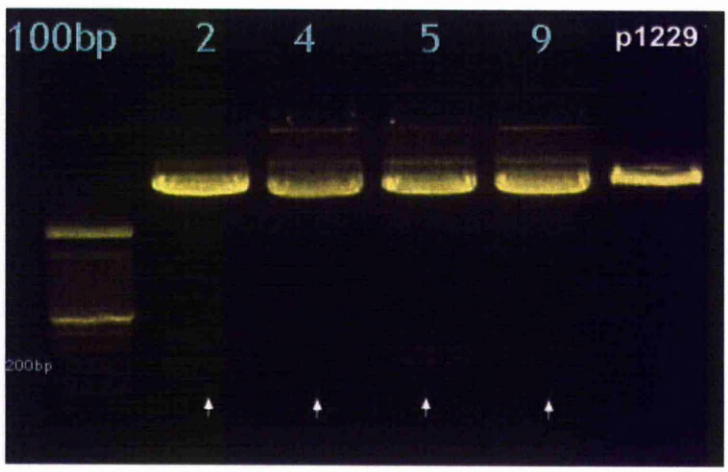


Figure 38 (A) STET prep output showing the presence of a band running at about 7kbp. Circular DNA can be adopt a number of conformations such as supercoiled where it will run faster through the gel than it it were an open circular plasmid. This band, which perhaps serendipitously runs at ~ 7kbp, indicates the possible presence of the putative ECR1-Hb9:p1229 construct.

(B) Double digestion of the Hb9:p1229 construct by NotI and SpeI showing the released 200bp fragment (arrowed) indicating that the enhancer had been cloned into p1229. (0.8% agarose gel).

Figure 39 Results from the sequencing from T7 direction of Hb9-ECR1:p1229 Key: Red, reverse-complement of reverse primer; Green. Reverse complement of forward primer; Turquoise, Reverse complement of ECR sequence, Restriction enzymes sites are annotated and also indicated specifically as follows: NotI (puce), EagI, (puce underlined), SpeI (teal).

3.4.5 Injection and transfection of the ECR1-Hb9:p1229 construct into chick tissue.

The techniques used for egg and embryo maintenance, injecting and electroporation were developed and refined over the course of the project. To develop the techniques for exposing of the embryo, electrode placement, needle shape, size and preparation, injection technique, electroporation parameters was varied to give the best results.

The electroporator that was used for this experiment was battery-powered and was rather primitive having only two controls; output voltage and pulse length. Each pulse was triggered manually by push-button and there was no current control. Many experiments were run under differing conditions before electroporation was achieved; the output voltage was altered from 15V through to 60V, pulse length changed from 50 msecs to 100 msecs and number of pulses was varied from 3 to 5 with the gap between pulses ranging from as fast as was manually possible to about one second. Electroporation was achieved at a setting of 40V upwards but survival was lower at the higher voltages and indeed this 40 volts and upwards contrasts with the 15V quoted in other papers (Itasaki et al., 1999, Croteau and

Kania, 2011, Islam et al., 2012). The best compromise between electroporation success and survival was found to be five pulses of 40V each, at random speed and with a pulse length 50 msecs. At these parameters between 20% and 50% of the embryos survived and showed green in the neural tube.

The optimum location of the electrodes in relation to the embryo was found to be important and relatively difficult to achieve due in part to the instability of the embryo and its substrate. When positioned on either side of the embryo both electrodes had to make good electrical contact with the egg yolk. This was done in the opening in the vitelline membrane and further assisted at times by the application of a few drops of L15 or PBS. The electrodes also had to be oriented so that the plasmid, as it was being attracted to the positive pole would travel towards the desired location. This meant attempting to ensure that the neural tube was centred between- and parallel to- the electrodes and in the correct plane. When positioning the electrodes extreme care had to be exercised to ensure all these parameters were achieved and that the embryo was not disturbed too much and also that contact with the extra-corporal main blood vessels was avoided.

The gap between the electrodes was ideally set between 3mm and 5mm depending on the size and orientation of the embryo. It appeared that at a narrower gap the focus of the electrical pulse was too narrow to achieve a good electroporation result. At a wider gap it was assumed the voltage drop was too great, also resulting in poor results.

The electrodes initially used were an in-house construction using platinum wire but these were later changed to commercially-made gold-plated electrodes.

The initial, test injections were made using with Fast Green dye dissolved in PBS and the GFP-reporter vector. Only when the electroporation was shown to function by the presence of GFP positive cells in the targeted area was the transgenic reporter vector (Hb9-ECR1:p1229) co-injected with the GFP.

Because the signal from the p1229 chromogenic vector is only expressed in response to specific signals within the cell and could only be observed following sacrifice of the embryo, a fluorochromic vector was co-injected with the p1229 reporter vector. This GFP-expressing produces a signal that is independent of intracellular signalling, is visible within hours, and being fluorescent, no processing of the subject is required.

3.4.6 Transfecting the chick embryos with the ECR1-Hb9:p1229 construct.

The contents of fertilized eggs of fifty-five hours incubation, were placed in individual cling-film hammocks and the embryos micro-injected with the plasmid mix (GFP - 2 μg/μl, ECR1-Hb9: p1229 – 3 μg/μl) and electroporated with 5 pulses of 40V at 50 msec pulse length and random intervals between pulses. Three embryos were successfully injected.

Following 44 hours in a passively humidified incubator at 37 °C two of the embryos were found to express fluorescence in the neural tube (Figure 40 A) and were processed as described in the Embryo Staining Protocol. Following 5 hours incubation at 37°C, blue staining, co-localised with the GFP fluorescence was observed (Figure 40,B).

The embryos were then processed as per "Harvesting Embryos" in Materials and Methods chapter, and 10 micron frozen sections were made. From these it can be seen that only one side of the neural tube is GFP-stained showing the integral negative control of this system (Figure 40 C) and that the staining occurs in the motor-column of the neural tube (Figure 40 D) matching the location in mouse (Figure 29).

3.4.7 Hb9 antibody staining to identify location of the Hb9 protein in Chick Neural Tube.

Hb9 antibody staining using MNR2 (University of Iowa) was found to stain the nuclei of the neural-tube motor–column (Figure 41). This was the same region that the ECR1-Hb9:p1229 reporter plasmid construct caused staining (Figure 40D). The negative control, using no 1° antibody is completely blank under fluorescent illumination (Figure 42).

The fact that staining occurred in the equivalent position in the chick to that previously demonstrated in the mouse (Figure 29) strongly indicated that this tool had functioned correctly in the chick embryo.

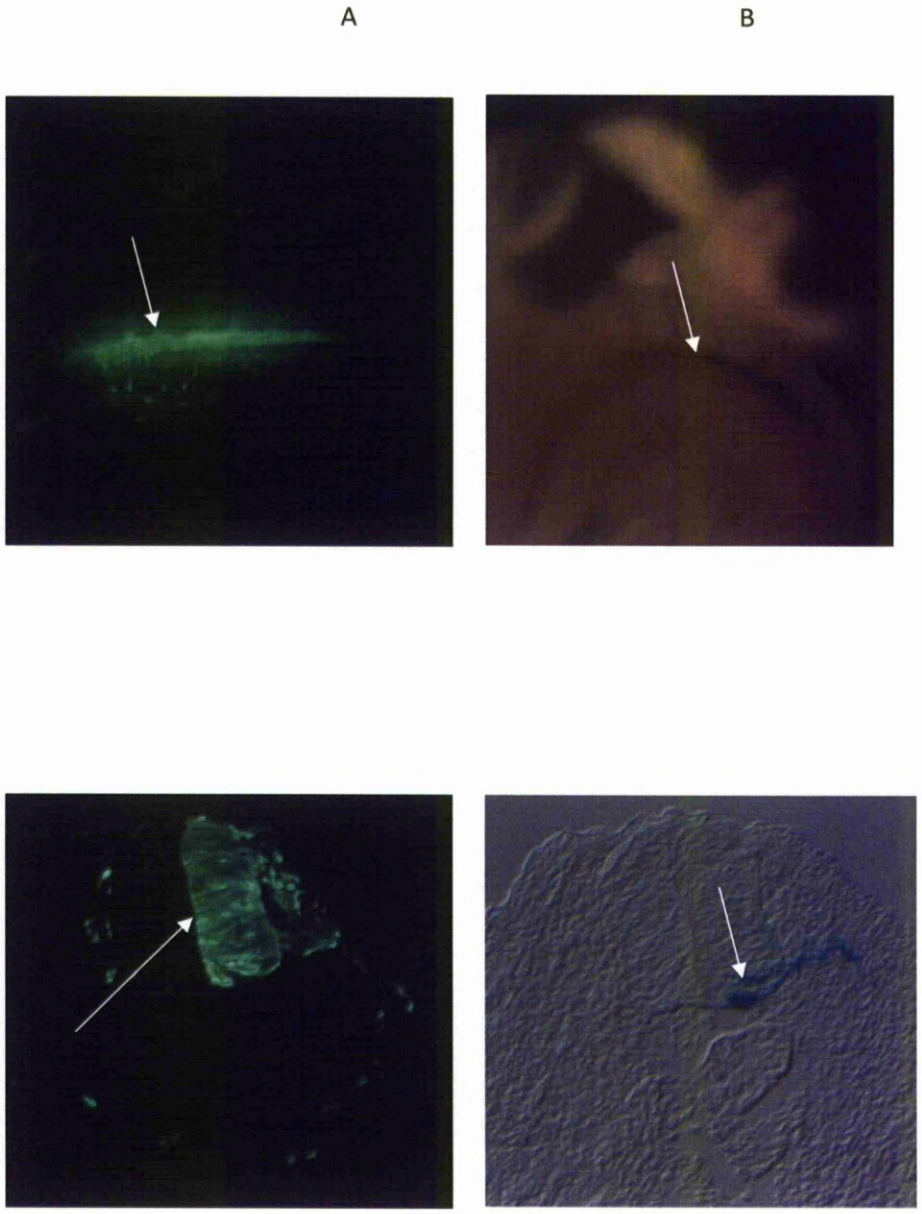


Figure 40 Whole mount (A,B) and 10 micron sections (C,D) of embryonic day 4 chicks showing fluorescence (A,C) indicating the location of the electroporation and Hb9-enhancer *LacZ*-driven blue staining (B,D) following successful micro-injection and electroporation of the neural tube.

Figure 41 Hb9 antibody staining is in the motor neuron column. (A) Embryonic day 5 chick neural tube (outer edge indicated with blue arrow) stained with MNR1 1° Antibody 81.5C10 @ 1:5 (University of Iowa) highlights the motor column (orange arrows). (B) DAPI staining of the same section indicating the location of the nuclei.

Figure 42 Negative control of MNR1 antibody in 10 micron section of torso of HH stage 25 chick. (A) Phase image showing neural tube, notochord and DRG, whilst (B) shows negative control fluorescent image of same section.

3.4.8 Investigation into Transcription Factor Binding Sites found within the ECR1 enhancer of Hb9.

The 212bp sequence that had been cloned into the p1229 reporter vector was scanned for TFBS using the ECR Browsers link to TRANSFAC professional V10.2 and rVista V2.0.This showed that of the 53 TFBSs in the base sequence (Human) and 49 in the second sequence (chick) 19 TFs were conserved and aligned between the species. This indicates that this group of TFs could be important in the regulation.

Which TFs are driving the reporter gene expression? This still appears to be a nascent area with a paucity of specific information on what TF or group of TFs drive Hb9 expression. GATA1 (a.k.a. NF-E1) is a transcription factor that binds the regulatory regions of many viral and cellular genes (Becker et al., 1994).

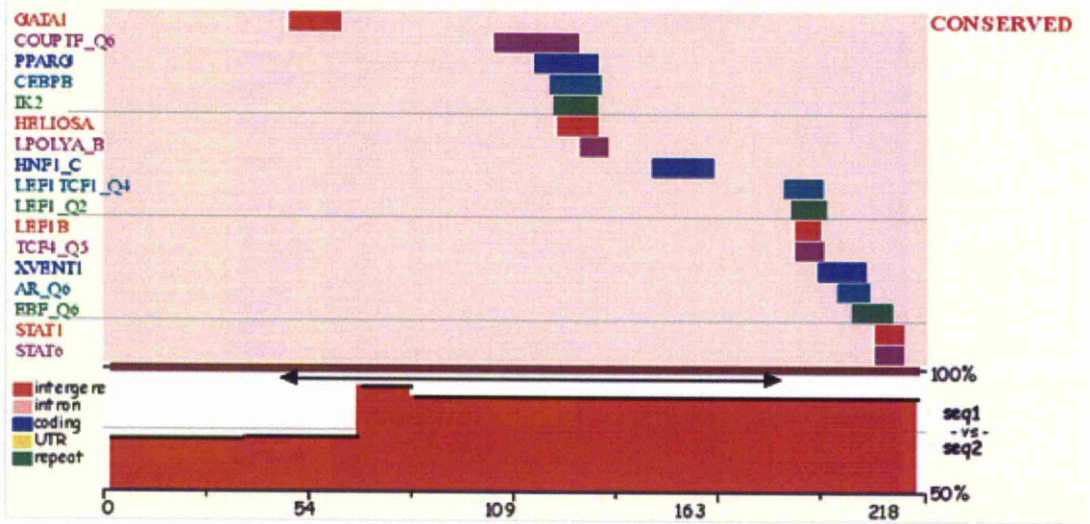


Figure 43. TFBS conserved and aligned between human and chick in the 212bp ECR upstream of Hb9(Mnx1) in chick that had been demonstrated to act as an enhancer. TFBS conserved between human and chick in the 125 bp enhancer region are indicated by arrowed line and appear to form three groups with GATA1 and HNF1_C forming groups on their own with the remaining TFBS in the other group. The region Hox/Pbx consensus sequence that Nakano disrupted and so destroy the enhancer effect is at the HNF1_C TFBS

3.5 Discussion

As it had been unknown if an enhancer from a chicken inserted into the p1229 vector would function in chick it was necessary to test that it.

The ECR Browser (Ovcharenko et al., 2004) was shown to be a very useful tool in identifying the corresponding ECRs between chick and mouse because by using the ECR Browser to compare these two genomes the 125bp region that had been identified (Nakano et al., 2005) as an enhancer of Hb9 in mouse was identified as an ECR in mouse and t the equivalent sequence was also shown to be an ECR in chick. The ECR grab function, a feature in earlier versions of the ECR Browser, was used to capture the DNA sequence, and flanking regions, of the chick ECR and from this generate suitable primers using the Primer3 program (Untergasser et al., 2007, Koressaar and Remm, 2007). Although difficulties were encountered in PCR-amplification of the ECR sequence from the whole-chick DNA the success at the lower T_A of 55°C and 55.3°C by gradient PCR would indicate that the convention of setting T_A at 2°C below T_M was ill-advised and that a greater temperature difference, such as a T_A 5°C below T_M would be more appropriate in this case. However care must be taken when lowering the T_A as this increases the likelihood of the primers annealing to the wrong target (Metzenberg, 2007). Although Primer3's method for calculating T_M is sophisticated, such as using the oligo melting temperature formula and the table of thermodynamic parameters given in (Breslauer et al., 1986), this is still only an approximation of the empirical situation and may vary in accuracy with different DNA and primer combinations.

As it was, the T_A in the gradient PCR run that produced the product had a T_A of 55.3°C putting it only 2.7°C below the lowest T_M of the two primers as calculated by Primer3. What is surprising is that two runs had previously been made at 55°C using a conventional PCR machine that resulted in no PCR product being produced. This may just random chance or might indicate a slight temperature differences between the two machines either of which could result in fewer primer-annealing events in the first few PCR iterations. Due to the exponential nature of PCR product formation small differences at the start of the cycle can result in huge differences after thirty iterations. This can result in sufficient PCR product to be easily observed for the reaction that had an adequate number of primer annealing events, and in the other case, where there was a low number of primer annealing events, too little PCR output to be observable.

The very bright GFP signal produced by the expression vector under its βActin promoter indicated that this functioned well in the chick neural tube. The in-built negative control was also shown to perform appropriately because only one hemi- section of the neural tube was found to GFP-stained indicating that the GFP-expression was not an artifact or of an endogenous nature.

The two major unknowns about how the p1229 reporter vector would function in chick were answered in this proof-of-concept process. These were whether p1229 would function when stimulated by binding of the correct TFs to the enhancer sequence or conversely would it be 'leaky' and drive the reporter gene without any stimulation. As previously discussed, some minimal promoters are 'leaky' in some tissues/organisms and so an appropriate minimal promoter must be selected. Both of these questions were answered during the work described in this chapter. The human βglobin promoter was indicated not to be leaky in two ways; firstly, empty plasmids that were transfected into chick neural tubes produced no signal and secondly and more telling, the reporter-vector-generated blue

staining was located solely to motor-neurons, matching the location in the mouse as anticipated, even though the whole hemi-section of the neural tube was transfected as evidenced by the co-transfected GFP-expressing plasmid.

Due to the success of the blue staining produced by the chick Hb9: p1229 ECR1-containing construct proved to be very effective because out of four GFP-expressing chick embryos three of these were found to have expressed βgal by showing blue stain upon processing.

Taken together, the results from this proof-of-concept showed that it was a robust positive control that had successfully demonstrated that the system functioned in chick.

This then served to instil sufficient confidence to proceed to the next phase in the process – testing the system against a putative enhancer from the *LSAMP* genomic region.

Chapter 4.

4 ECR adjacent to the *LSAMP* gene shown to possess enhancer functionality.

As previously described, some non-coding sequences that have been conserved over evolutionary time have been shown to act as enhancers of gene expression (Prabhakar et al., 2006). These ECRs can be exposed by comparing the genomes of different species and identifying the non-coding conserved regions. By cloning these sequences into a suitable reporter plasmid and transfecting appropriate cells with this reporter construct, the reporter vector's signal will be expressed when the ECR sequence has acted as an enhancer.

It was demonstrated in the previous chapter how the Hb9-enhancer p1229-based reporter construct (ECR1-Hb9:p1229) functioned successfully in chick, by driving expression of the reporter gene *LacZ* in appropriate locations in the chick motor column. The next step in the process was to use the p1229 reporter plasmid to investigate if ECRs adjacent to, and within the *LSAMP* transcription unit, potentially act as enhancers. LAMP, a cell adhesion molecule of the IgLON family was selected for investigation due to its possible implication in anxiety (Nelovkov et al., 2006, Catania et al., 2008). *LSAMP* is a longer and more complex gene than *Hb9*. Because of the larger genomic extent of *LSAMP* the possibilities of where the controlling sequences may reside also increases as does the possibility that ECRs associated with other genes may be intermingled with those of *LSAMP*. Because *LSAMP* has a number of isoforms and is active in various tissues at different stages of development the exquisite control required to ensure its correct temporal and spatial expression are probably very intricate requiring many enhancer and repressor sequences. The search for enhancer regions in and around *LSAMP* is by its nature potentially more complex than searching for enhancers in a smaller, simpler, gene.

4.1 AIM

The aim of this chapter is to select ECRs for testing as putative enhancers; to amplify these sequences and insert into a reporter plasmid; to inject these plasmids into the neural tube and brain of embryonic chickens and electroporate to generate transfected cells and to test if the ECR sequences act as enhancers by driving expression of *βGal/LacZ* within the reporter plasmid construct.

4.2 METHODS

4.2.1 Selection of first three ECRs in the genomic region of LSAMP in chick.

Following the positive results obtained from the Hb9 enhancer proof-of-concept, the region upstream of the short *LSAMP* genomic sequence, Figure 15, was scanned for ECRs by comparing chick (target) against human (base) using the ECR Browser (Ovcharenko et al., 2004).

LSAMP contains many sequences designated as ECRs, for example between the distal and proximal TSSs (Table 11) it can be seen that there are 113 at stringency 100bps/70%. What is designated as an ECRs varies with the stringency employed as can also be seen in this table.

Length specified/bp	Percentage similarity	BASE:Chicken Target: Human	BASE:Human Target: Chicken
100	85	39	46
100	70	113	127
350	70	23	28

Table 12 Number of ECRs in *LSAMP* between proximal and distal TSSs at differing stringency. ECR Browser

Due to the large number of ECRs a decision was made to start with the first three ECRs upstream of the proximal *LSAMP* TSS to see if the methodology would work prior to testing many more. These three ECRs were exposed at a stringency of 100bps/80% identity by the ECR Browser, this value giving a reasonable number of persistent ECRs The nearest ECR to the proximal TSS site, ECR1, was 29 kbp upstream of the TSS, in chick, with ECRs 2 and 3 further upstream (Figure 44, Figure 45). The Hb9 ECR had been much closer to the TSS i.e. 1.7 kbp upstream in chick, 11 kbp in mouse(mu8 database build) and ~7.7 kbp(mu9 database build) the 29 kbp found for *LSAMP* was well within the 60kbps distance at which the known human β-globin enhancers act(Porcu et al., 1997). ECR1 and ECR3 were still being exposed at 95% and 100bp, possibly due to its shorter length ECR2 disappeared at 90% and 100bp.

4.2.2 Design of primers to allow for directional cloning

The ECR sequences plus flanking regions of 100bps were obtained from the ECR Browser and using Primer3 suitable primers were obtained for ECR1, ECR2 and ECR3 (Table 23, Appendix). To allow for directional cloning restriction sites suitable for cutting by NotI (GC^{V}GGCCGC) and SpeI (A^{V}CTAGT) restriction enzymes were appended . Because of the proximity of the restriction enzyme cleavage sites (marked by ^{V}) to the ends of the DNA sequences an extra four nucleotides, complementary to the DNA, were appended to the 5' end of the primers to allow for efficient restriction enzyme cutting. These extra nucleotides were selected to bind to the DNA sequences where possible but in the ECR2 reverse primer this would have resulted in a hairpin structure and the sequence was modified to avoid this (Table 23).

The DNA of all three ECRs was amplified by PCR as previously described in Materials and Methods, with T_A=58°C. Following DNA purification by phase-lock gel the PCR product was double digested with NotI and SpeI restriction enzymes (Fermentas FastDigest). The digested

products were isolated by 1.8% gel electrophoresis and visualised by SyBR green fluorescent nucleic stain and found to be the expected sizes (Figure 47). The PCR product was then excised from the gel, glass-wool fractured, ethanol precipitated and re-suspended in water. Ligation into the p1229 reporter vector that had also been double digested with NotI and SpeI (Figure 48) resulted in the desired ECRx-*LSAMP*:p1229 constructs (where x = 1,2 or 3). These were transformed into *E. coli* TOP10F' (ECR1-*LSAMP*:p1229) or DH5α (ECR2-*LSAMP*:p1229, ECR3-*LSAMP*:p1229) and the population expanded overnight and initial selection was made by spread plate. As blue/white screening is not supported by the p1229 plasmid, colonies were chosen by their size, isolation from neighbours and their lack of satellite colonies. A scraping of the selected colonies was PCR-screened using the primers originally used to amplify the strands. The PCR output was then run through gel electrophoresis and a number of colonies from each ECRx-*LSAMP*:p1229 construct were then selected, based on the appropriate molecular weight and the intensity of SyBR green signal from the band. Selected colonies were then expanded overnight and the plasmids extracted using a plasmid prep kit. Clones for plasmids containing ECR1 (clones R5 and ABM6), ECR2 (clones 5 and 8), ECR3 (clones F and I) were sequenced from the direction of both the T7 and T3 primers within the p1229 plasmid. All of the sequenced ECRx-*LSAMP*:p1229 plasmids were found to have the correct ECR inserted in the desired direction, i.e. from T7 promoter (Figure 77, Figure 78). The results obtained from the sequencing were of high quality, with only one nucleotide difference in the ECR sequences. If a higher fidelity had been required multiple sequencing passes would have been ordered. Many researchers have three sequencing runs as the default. The sequencing vendors also offered higher fidelity sequencing, at extra cost, but probably because of the short lengths of DNA involved, the basic service was found to be adequate. Sequencing technologies have continued to improve (Hall, 2007) and these massively parallel methods offer a high quality service at an affordable price.

The transformed bacteria for each of the constructs were expanded overnight in LB broth with 50 μg/ μl ampicillin at 37°C in a shaking incubator and endotoxin-free plasmid-preps were made. As a final concentration of 2 - 6ug/μl was specified in the electroporation protocol the ECRx-*LSAMP*:p1229 constructs were concentrated by ethanol precipitation and re-dissolved in PBS plus 1mM Mg^{2+}. They were then assayed to check concentration and λ260/280μm ratio (Table 12).

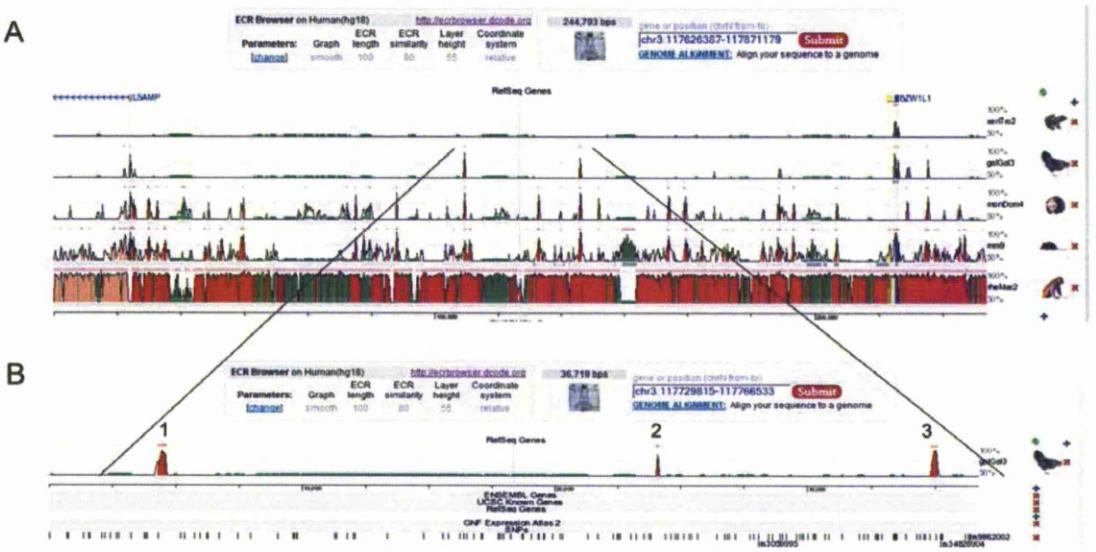


Figure 44 First 3 *LSAMP* ECRs. ECR Browser output for the first three *LSAMP* ECRs, showing (A) ECRs in the region upstream from Human *LSAMP* (RefSeq entry); Comparison made using Human as base organism and macaque, mouse, opossum, chick and frog as target organisms, using a stringency of 100bp/80%. As can be seen sequence similarity decreases with evolutionary distance. (B) Enlarged section showing the three ECRs in greater detail. The three selected ECRs are marked 1,2 and 3.

NOTE1: Due to an artifact of the system, although the pink ECR marker does appear, the vertical conservation plot for ECR2 does not appear in (A), but only emerges upon zooming-in as shown in (B).
NOTE2: The gene BZWILI, official name BZW1P2 basic leucine zipper and W2 domains 1 pseudogene 2, is a pseudogene and therefore does not code for a protein (Muzny et al., 2006).

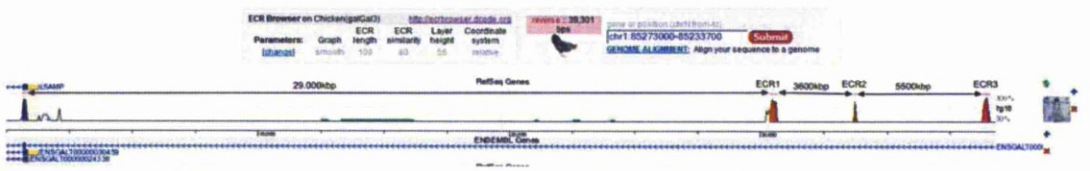


Figure 45 First three ECRs in chick. ECR Browser output showing (orientation flipped due to *LSAMP* being insertion in opposite direction between chicken and human), ECRs in the region upstream from chick *LSAMP* (RefSeq entry); Chick was used as the base organism and human as target organism. ECRs are marked 1,2 and 3 and were identified at 100bp/80% stringency. The distance (bp) that ECR 1 is upstream of *LSAMP* and the gap between other ECRs are shown. The short *LSAMP* isoforms g11/g19 are ~286,000 bp – see Figure 46

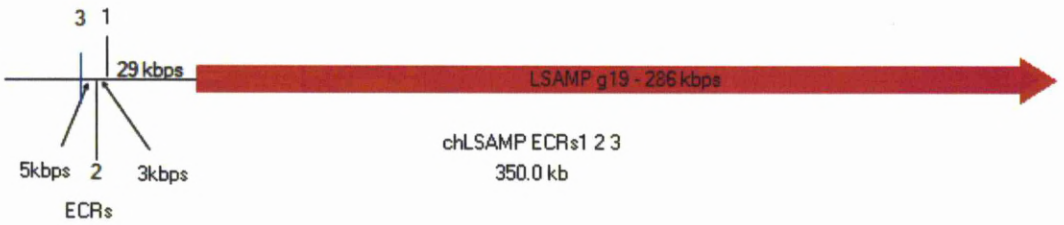


Figure 46 Distance map of ECRs from g11/g19 *LSAMP* isoform

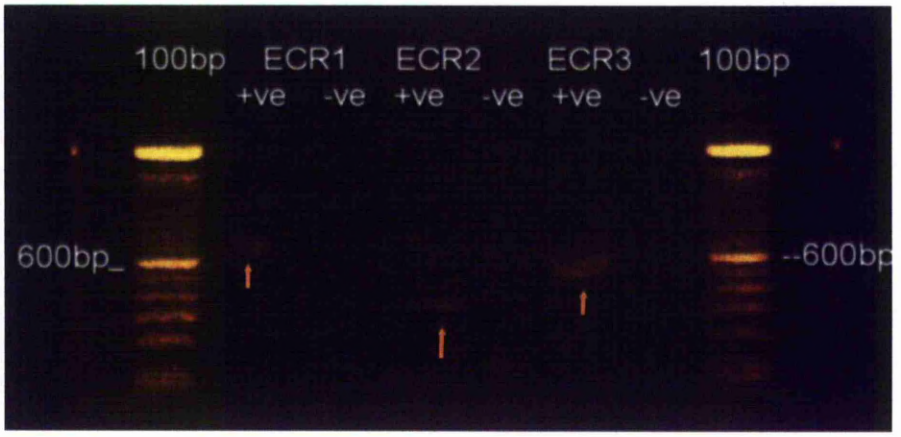


Figure 47 Amplification of ECR1, ECR2 and ECR3 from whole genomic Chick DNA. 1.8% agarose gel showing the (arrowed) amplified DNA fragments for ECR1 (expected size 500bp), ECR2 (expected size (200 bp), ECR3 (expected size 380bp).The negative control track (-ve) is empty for all ECRs. The 600bp line is identified on molecular weight marker track, each band below this is separated by 100bp. SyBR green was mixed with all samples prior to loading on gel resulting in the fragments running slower in the gel than they would without the SyBR green.

Figure 48 p1229 showing size shift. p1229 double digested with NotI and SpeI (track marked 'Digested') showing size shift against undigested p1229 (track marked 'Undigested'). This size shift indicates that at least one of the restriction enzymes had functioned correctly.

Plasmid Name	Concentration (µg/µl)	Purity (λ260/ λ 280 ratio)
ECR1-*LSAMP*:p1229	6.7	1.62
ECR2-*LSAMP*:p1229	11	1.63
ECR3-*LSAMP*:p1229	11	1.62

Table 13 Concentration and purity of the ECRx-*LSAMP*:p1229 transgenic constructs. A minimum (λ260/ λ 280 ratio) of 1.6 is required

4.3 Injecting and transfection

4.3.1 Plasmid mix preparation, injection and electroporation.

The plasmid was mixed with the eGFP vector and Fast Green dye as previously described to give a final ECRx-*LSAMP*:p1229 of 2 - 6 µg/µl. and NT of embryos were injected and transfected by electroporation (eight fast, manual pulses of 60V with electrode gap of 3-4 mm). The final concentration of the ECR and GFP-containing plasmid mixes varied in different runs (Table 14) in the transfection procedures run for this chapter.

ECR	Range of ECR Plasmid final [µg/µl]	Ratio ECR:GFP plasmids range
1	1.7 – 10.0	1.3:1 – 4.5:1
2	3.7 – 4.5	1.9:1 – 7.5:1
3 **Blue staining at**	1.9 – 7.3 **4.4**	1.8:1 – 10:1 **2:1**

Table 14 The range of the final concentration of p1229 plasmid containing ECR1, ECR2 and ECR3 used in the various transfections. The ratio of the ECR-containing plasmid and the GFP-expressing plasmid also shown. Blue staining was achieved at a final concentration of 4.4 µg/µl of the ECR1-containing plasmid at a ratio of 2:1 with the GFP-containing plasmid.

4.4 RESULTS

Following incubation, embryos which were successfully electroporated as evidenced by GFP staining (n=66) were processed to develop the chromogenic staining (Table 13). GFP staining of NT and co-localised *β-gal* (*LacZ*)-derived blue staining were observed in whole-mount embryos (Figure 49) and NT sections (Figure 50) for ECR3-*LSAMP*:p1229 (n=10). As with Hb9, only a subset of cells expressing GFP also stained blue (Figure 49) confirming that the human β-globin promoter was not leaky. The inbuilt negative control for electroporation is evident in that only one side of the NT exhibits fluorescence (Figure 50) and that the staining is as a result of the plasmids and not due to endogenous expression. While 29 embryos were successfully electroporated with ECR1-containing plasmid, as evidenced by GFP-staining, no blue staining was observed and ECR2-*LSAMP* testing was abandoned as no embryos were successfully electroporated.

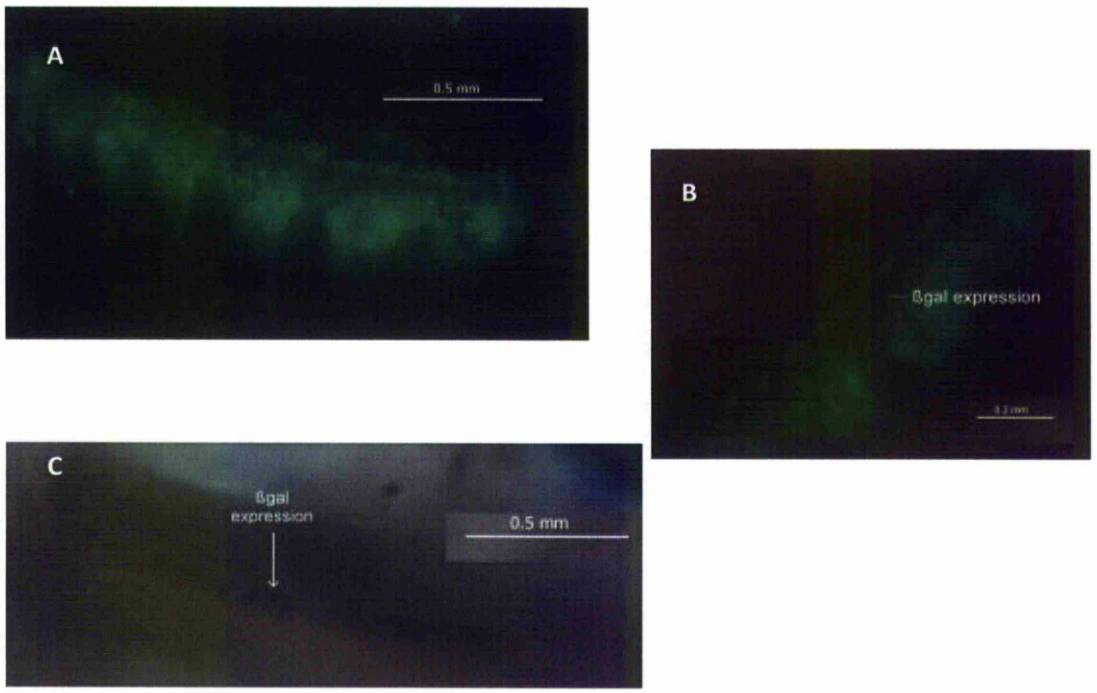


Figure 49 Whole mount E4 chick embryos showing (A) GFP expression. (B) Composite image of GFP and βgal expression as a result of ECR3, and (C) βgal expression as a result of ECR3 producing the blue staining of a sub-population cells in the neural tube.

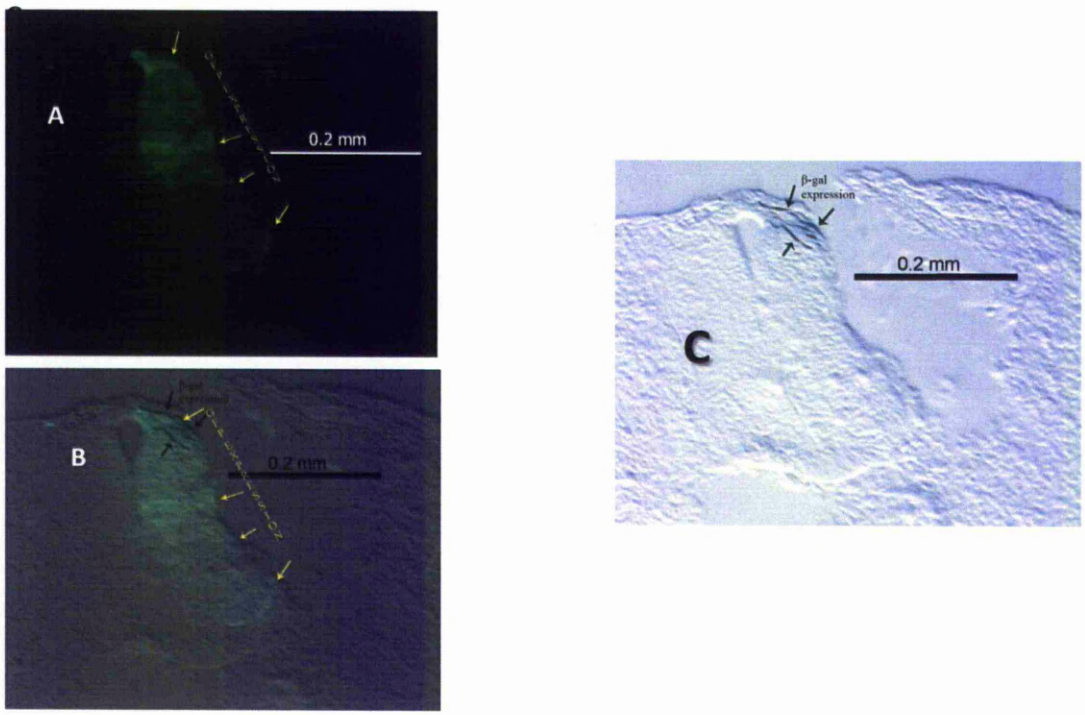


Figure 50 Sections (10 microns) of E4 chick NT showing GFP staining (A) composite image showing GFP and βgal co-localised (B) βgal expression (C).

ECR3 was shown to be an enhancer for *LSAMP* with 27% of the embryos that expressed GFP also expressing *β-gal (LacZ)* (Table 13).

The percentage of GFP-expressing chicks also expressing βgal increased with post-injection incubation time, increasing from 22% at 48 hours, through 33% at 72 hours and to 40% at 96 hours post-injection, although too few embryos were available at 96 hours for the result to be significant (Table 13). When post-injection incubation was allowed to proceed for 96 hours, dating the embryo as E6, the blue staining became more widespread. Both these results are consistent with the possibility that the expression of *LSAMP* increases under the influence of this enhancer and becomes more extensive in the neural tube as development progresses, this may be concomitant with an accumulation of *βgal* (Figure 52).

Conversely none of the GFP-expressing embryos transfected with ECR1-*LSAMP*:p1229 (n=29) produced any blue-stained cells indicating that ECR1 may be a repressor or that it was not activated at that stage of development, or that it does not possess a regulatory function. If ECR1 was an active enhancer it would be expected that about 7 of the ECR1-transfected embryos would also express *β-gal* based on the GFP- to *βgal* ratio of the ECR3-containing plasmid where 10 out of 37 were blue.

4.4.1 Transfection of chick Brain tissue using ECR3-LSAMP:p1229

LSAMP was reported to be expressed in the anterior midbrain from stage 14 (Kimura et al., 2001) so having established that this system functioned in the neural tube and was able to detect and report on the activation of enhancers an investigation into the brain was made to see if the NT success would be repeated.

Forebrain, Midbrain and Hindbrain regions were injected with ECR3-*LSAMP*:p1229 at HH stages 10-12 (n=10). The GFP-staining in the brain was found to be positive and widespread indicating that the micro-injection and electroporation methodology functioned efficiently and that investigation of enhancer activity in the brain would be feasible using this system. No blue staining, however, was observed in the brain (Figure 53).

The lack of *β-gal*-derived blue staining may either be due to lack of activation of the enhancer in the brain tissue at this developmental stage or too low a concentration of the ECR3-*LSAMP*:p1229 construct due to the large lumen in the brain at this age. To ameliorate this, the plasmid-concentration was increased and the bolus was deposited immediately adjacent to the target tissue but to no avail so ECR3-*LSAMP* may not act as an enhancer at this stage or in this tissue.

ECR	Location of Injection	H.H. Stage when Injected	Hours incubated post-injection	H.H. Stage when Harvested	GFP +ve embryos	βGal +ve embryos	% blue embryos of those that were GFP+ve
ECR3	NT	14	48	23	23	5	22
			72	25	9	3	33
			96	28	5	2	40
			Both 72+96	25 + 28	14	5	36
Total	NT				37	10	27
ECR3	B	11	48	23	4	0	0
			72	25	6	0	0
Total	B				10	0	0
ECR1	NT	14	48	23	7	0	0
			72	25	22	0	0
Total	NT				29	0	0

Table 15 Summary of the results of the embryos injected with ECR1-*LSAMP*:p1229 and ECR3-*LSAMP*:p1229 reporter indicating the location (NT – neural tube, B – brain) the Hamburger-Hamilton stage, the number of embryos expressing GFP and βgal and percentage of the GFP-expressing embryos that also expressed βgal resulting in blue staining.

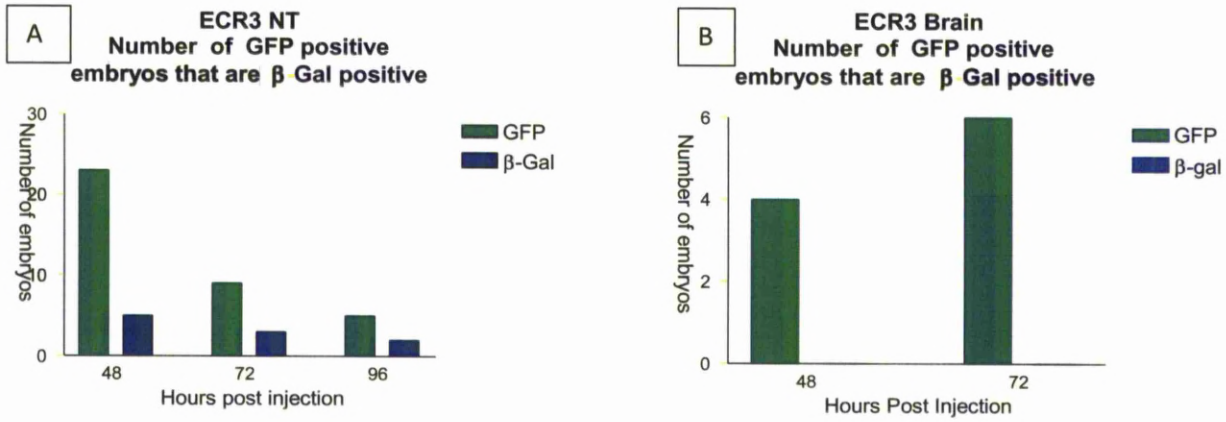


Figure 51 (A) Expression of GFP and ECR3-LSAMP-derived β-gal staining in chick neural tube by post-injection time and (B) GFP staining in brain.

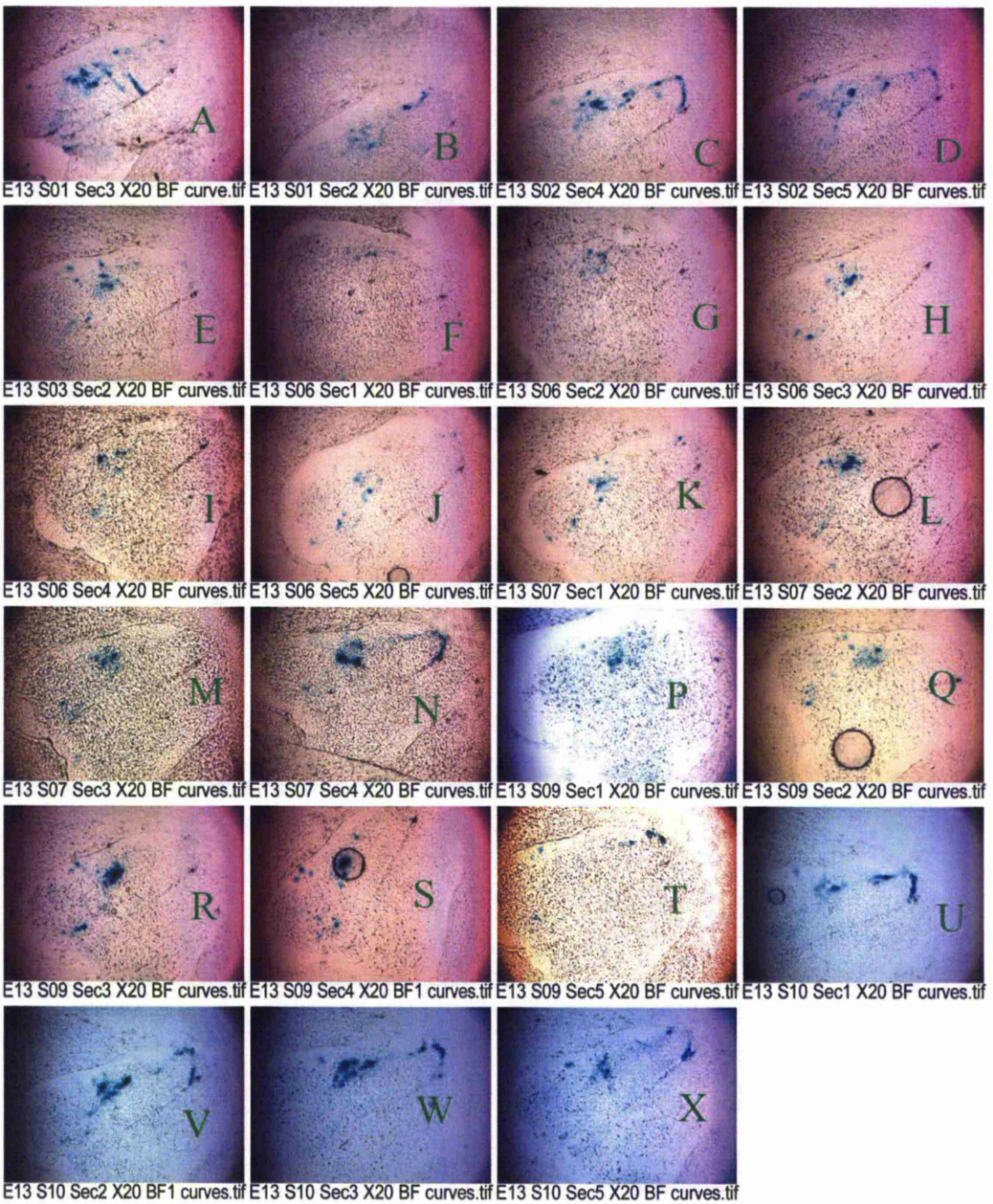


Figure 52 ECR3-driven blue staining at 96 hours post injection in sequence of 10 µm section taken from the neural tube in an anterior to caudal direction. Sections are bright field view of HH Stage 28 embryo NT (dorsal at top right) transfected with ECR3-containing p1229 reporter vector showing extensive blue staining 96 hours incubation post injection.

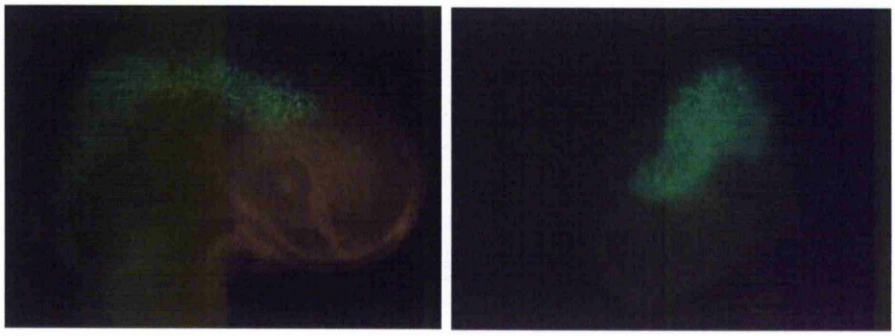


Figure 53 GFP expression in brain region of chick embryos at 48 hours post-injection (HH stage 23). Image of HH stage 23 injected at HH stage 10 in midbrain hindbrain region with ECR3-containing p1229 plasmid co-injected with GFP-expressing plasmid. No blue staining was visible following processing.

LAMP antibody staining was undertaken to show the location of LAMP expression. If, as expected, ECR3-*LSAMP* drives LAMP expression the blue staining should be co-located with the antibody-driven staining.

LAMP antibody-staining of an E6 chick's neural tube showed the LAMP-expressing region to be in the region of the commissural axons and not so highly expressed where the βgal expression had been indicated Figure 54. The β-gal is expressed in the soma of the neurons and the LAMP expressing axons of these developing neurons come into close proximity at the commissure as would growth cones of developing axons at the moment resulting in this obvious antibody staining. The fact that LAMP is a membrane protein may explain the pattern of staining due to the relative amount of membrane in the axons compared to the lower amount in the soma.

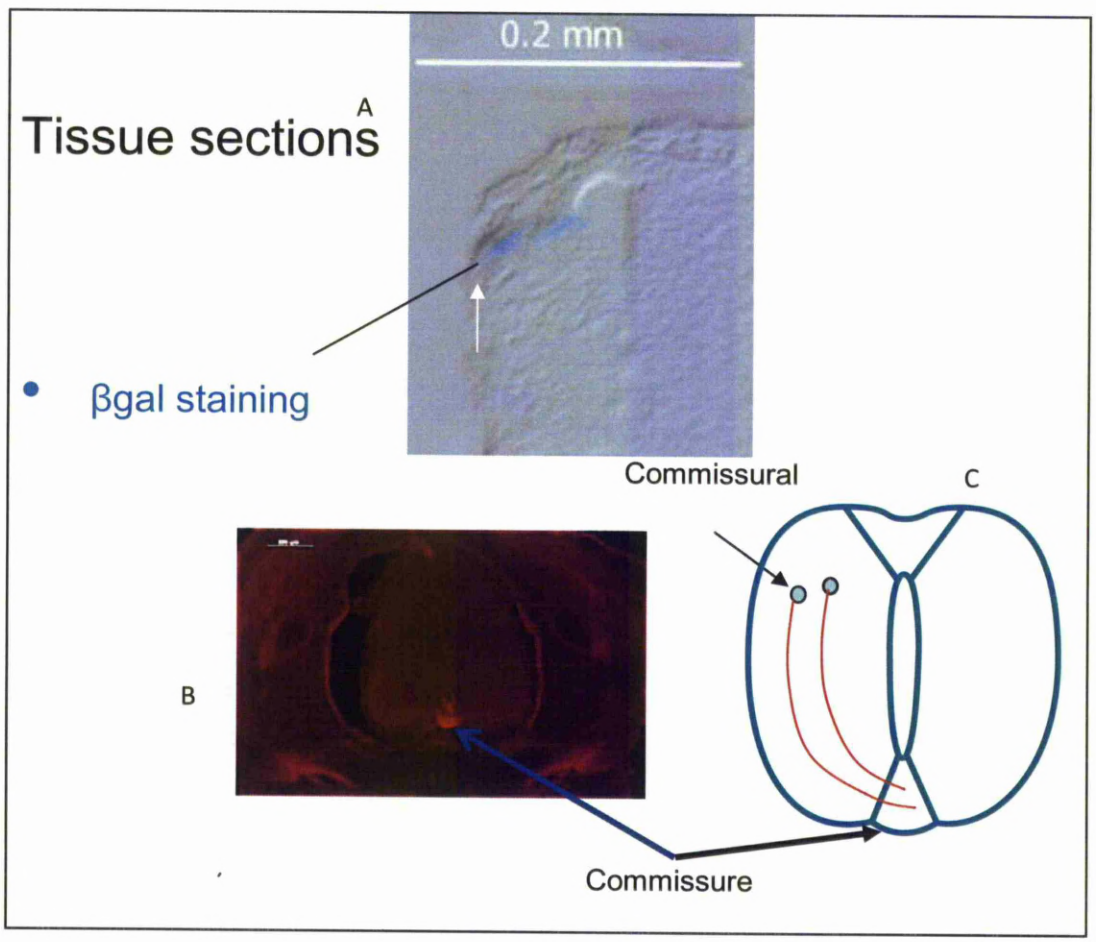


Figure 54 E4 chick NT sections. (A) ECR3-*LSAMP*-driven βgal staining (B) *LSAMP* antibody staining very evident at commissure.(C) Diagram of neural tube indicating cell bodies of commissural neurons and their axons.

4.5 Transcription factor binding sites located within ECR1, ECR2 and ECR3 of the *LSAMP* gene.

Upon analysis of the transcription factor binding sites within the ECRs it was found that ECR2 had only four conserved TFBSs in 2 groups (Figure 55) of which only two (PAX3, MEIS1-using simplified names), were in the FANTOM4 database. No interactions between them were indicated.

Within the ECR1-sequence there were 32 conserved TFBSs in about 9 groups (Figure 56) and within the ECR3-sequence there were 33 conserved TFBSs in about 10 groups (Figure 57) indicating that ECR1 is as complex as ECR3 and some effect could be expected from it. As many of these TFBSs share much of the same motif it is probable that the number of groups is of higher significance than number of individual TFBSs. With ECR1 it is interesting to note that there are two TBP (TATA box binding protein) binding sequences. TBP is usually found in the promoter region 25bp upstream of the TSS.

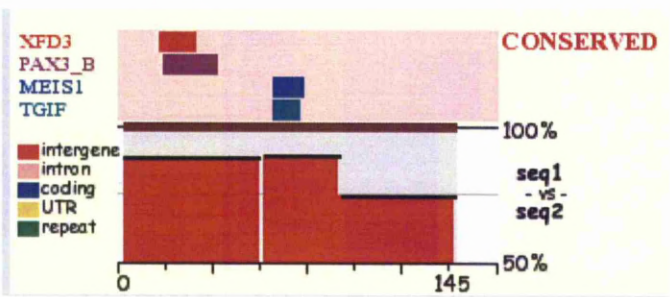


Figure 55 ECR2-*LSAMP*: 4 conserved TFBSs showing their relative positions, overlapping to form 2 groups. Base organism Human, target organism Chick. (TRANSFAC Professional V10.2 library, analysed by rVISTA V2.0 (Loots and Ovcharenko, 2004) accessed through ECR Browser)

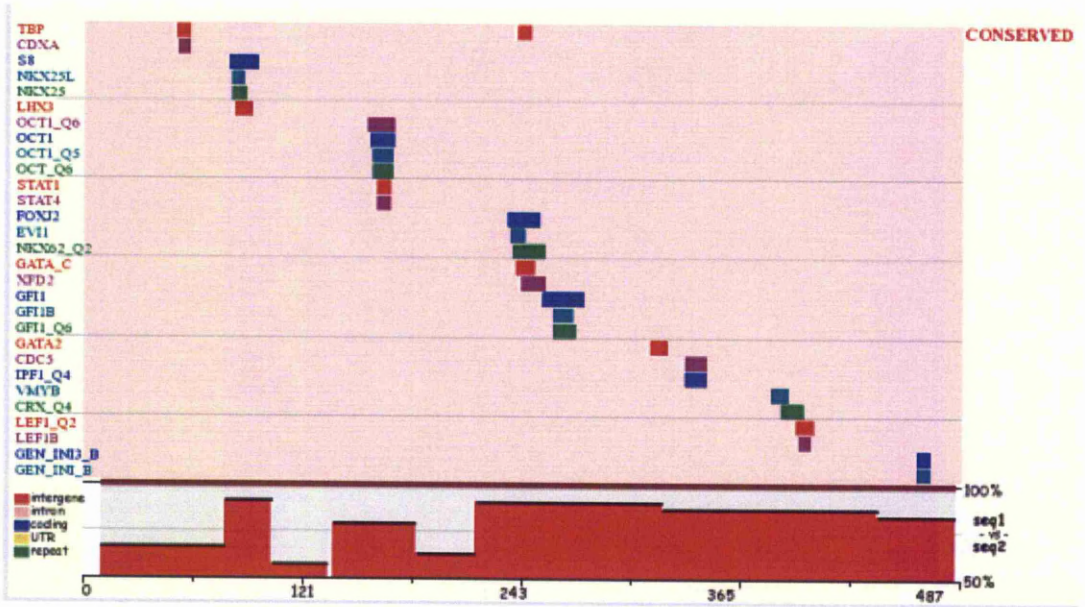


Figure 56 ECR1-*LSAMP*: 32 conserved TFBSs showing their relative positions and overlapping to produce ~9 groups. Base organism Human, target organism Chick. (TRANSFAC Professional V10.2 library, analysed by rVISTA V2.0 (Loots and Ovcharenko, 2004) accessed through ECR Browser)

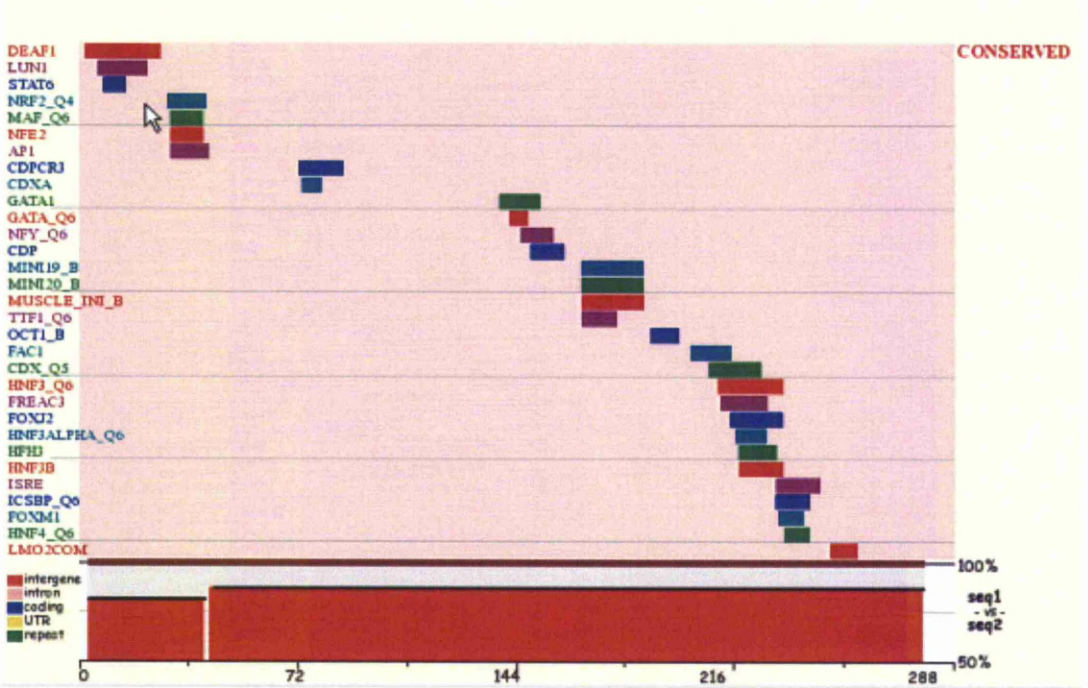


Figure 57 ECR3-*LSAMP* : 33 conserved TFBS showing their relative positions and overlapping to form ~10 groups. Base organism Human, target organism Chick. (TRANSFAC Professional V10.2 library, analysed by rVISTA V2.0 (Loots and Ovcharenko, 2004) accessed through ECR Browser)

4.6 Discussion

During the work done to complete this chapter the embryo-handling, injecting and electroporation methodologies were refined and improved. Although only a 21% success rate of injections (n=465) to GFP-expressing embryos (n=99) using the current battery-powered electroporator was achieved it was being proposed that a newer electroporator should be obtained which would hopefully provide better results. The current electroporator had no pulse rate facility or over-current protection and the voltage setting (40V-60V) that we had to use to achieve transfection was higher than the norm of 20V-25V (Itasaki et al., 1999) possibly leading to damage of the embryo.

This system was shown to be able to identify ECR3-*LSAMP* as an enhancer of gene expression and to differentiate between a functional enhancer and ECR1-*LSAMP* which was not shown to be an enhancer and at this stage in development in this particular tissue.

Of interest was the fact that only some of the embryos that had been successfully electroporated (as indicated by the GFP staining) also stained blue. It is unlikely that this difference was a result of differential expression of TFs by the embryos because others processed at the same stage of development stained blue. If only a subset of the neurons activated the enhancer and none of these particular cells were successfully electroporated in the βGal-negative cells, this may explain the observed result. An alternative explanation might be that blue was expressed but at such a low level that it was not observed, a fluorescent reporter may prove more detectable.

The final concentration of the ECR-containing plasmid nor the ratio of the ECR and GFP-containing plasmids did not seem to be a contributory factor as the ranges tested

overlapped that of the transfection that resulted in the ECR3-containing plasmid producing blue embryos (Table 14).

A variable not fully catered for here is the amount of plasmid that was injected and/or remained in the neural tube at the time of electroporation. As the neural tube closes up in line with increasing development age it was attempted to ensure that all the embryos were at the same HH stage at time of injection. As it was impossible to obtain eggs that had been fertilized and held in exactly the same conditions until injection, this was impossible to achieve and so embryos at a range of development stages were found in each batch of injections. A further complication was that the variations in development stages became more pronounced the longer the fertilized eggs were held in cold-storage prior to injection. Another issue was that, at times, there was leakage of the plasmid some of the injected bolus naturally leaked from the neural tube, and although speedy working helped ameliorate this. Taken together, the size of the neural-tube's lumen and the amount of leakage would have caused variations in the amount of plasmid available for the electroporation stage.

Evidence would indicate that ECR3-*LSAMP* acts as an enhancer of *LSAMP* in the neural tube of E4 to E6 chicks, but no indication that it was acting as an enhancer in the brain at E4/E5 was observed. There are several possible reasons for this such as the avian equivalent of the limbic-system had not yet developed or that expression of the reporter gene was insufficient to produce observable staining.

There is no evidence to support ECR1-*LSAMP* as an enhancer at this stage of development in the neural tube. Further testing for repressor function would be informative.

Although ECR2-*LSAMP* had been tested on only a few embryos (n=30), no GFP-positive results were obtained probably due to a failure of electroporation.

Upon analysis of the transcription factor binding sites within the ECRs it was found that ECR2-*LSAMP* had only four conserved TFBSs in 2 groups (Figure 55). Because TFBSs and their associated TFs work in a co-operative manner, ECR2-*LSAMP* may not be a very significant moderator of gene expression on its own and probably should not have been be tested in isolation i.e. the plasmid should have incorporated at least one of the other ECRs. Investigation into ECR2-*LSAMP* this was dropped as the number of TFBS was so low.

Further testing of ECR2-*LSAMP* in conjunction with other ECRs may indicate if it possesses a moderating function, but this was not undertaken. It may be, for example, that ECR2-*LSAMP* is a repressor of ECR3-*LSAMP* and if both ECR2-*LSAMP* and ECR3-*LSAMP* had been incorporated into the reporter construct the response from ECR3-*LSAMP* may have been reduced.

Why ECR3-*LSAMP* drove expression and ECR1-*LSAMP* did not is unknown. Naturally it is the population of TFs that populate the TFBSs that ultimately decide if a sequence has enhancer or indeed repressor functionality. ECRs 1 and 3 both had TFBSs for 9 and 10 groups respectively whereas ECR2-*LSAMP* had binding sites for 2 groups. This may mean that ECR2-*LSAMP* has very specific controls or that its effect is minimal. However, this does not explain why ECR3-*LSAMP* functioned as an enhancer and ECR1-*LSAMP* did not which must be related to the TFBSs in the ECR and the specific population of TFs being expressed at this stage by the cell.

While ECR3-*LSAMP* has been demonstrated to be an enhancer the function of the other two ECRs is as yet unknown. This system does not identify repressor function and so it is possible that ECR1-*LSAMP* and ECR2-*LSAMP* are repressors, or not active at this stage of development or have no regulatory function.

Although the evidence indicates that enhancers are orientation independent these three ECRs were inserted into the into the reporter vector in the same orientation as they are in the genome. Orientation independence was not tested for these three ECRs however results for the ECRs subsequently selected (Chapter 5) support the argument that enhancers are indeed orientation independent.

The results presented in this chapter show that the system was suitable for analysing ECRs for enhancer function as indicated by the results obtained for ECR3-*LSAMP*.

Chapter 5

5 Selection of additional *LSAMP* ECRs

The previous chapter described how the ECR designated as ECR3-*LSAMP* was shown to drive βgal expression. Having established that the tool worked, further ECRs needed to be investigated for enhancer activity in order to establish if ECRs could be rapidly screened using the chicken as a model.

Enhancers of gene expression are understood to be location independent (Banerji et al., 1981) however it would seem to be reasonable to expect that at least some of the control would be close to the target regions. For selection of the next tranche of ECRs the genomic region was divided into three regions; upstream of the g9 TSS to next gene IGSF11 (Group 1), between the g9 and g11/g19 TSS (Group 2), within the g11/g19 transcription unit (Group 3). The three original ECRs (ECR1-*LSAMP*, ECR2-*LSAMP* and ECR3-*LSAMP*) were located upstream of the g11/g19 TSS and so form part of the new 'Group 2' (Table 14,Figure 58).

5.1 *LSAMP* Expression profile

As indicated in Chapter 1 , during early embryogenesis LAMP expression was found to be active by HH stage 10 in the anterior region of the neural tube However at stage 14, the expression was limited to most of the neural crest cells, the anterior midbrain region and the notochord. The notochord expression was absent at stage 25 but expression had moved to the floor plate and the lumbar region of the motor neuron columns. At the later stages of development, expression in motoneuron columns was seen at all levels of the spinal cord (Kimura et al., 2001).

5.2 AIM

The aim of this chapter is to continue to test the ability of this model to rapidly screen for enhancer activity in ECRs in the *LSAMP* gene's genomic region.

5.3 METHODS

5.3.1 *Selection of the additional ECRs in the chick LSAMP's genomic region.*

Initial identification of suitable ECRs was based on obtaining sufficient ECRs, using the ECR Browser, to provide a representative sample from these three regions by varying the stringency of selection (Table 14). The 16 ECRs thus selected were reduced to 8 by selecting those ECRs that were the longest and contained the highest number of discrete groups of TFBSs whilst ensuring there was representation from each group (Table 14). This resulted in the following list of ECRs:Group 1 – G1E2, G1E3;Group 2 – G2E4,G2E5,G2E6; Group 3 – G3E3,G3E4,G3E5 (Figure 58,Table 15). The numbering convention for the ECRs within the groups used the proximal TSS as the reference point with the ECR number increasing from this point, upstream and downstream.

The ECR regions were expanded by PCR from whole-chick DNA as previously described in Materials and Methods, and DNA regions were TA cloned into pCRII, expanded overnight and screened for putative enhancer sequence. Positive clones were expanded and digested to release the DNA fragment which was then ligated into p1229.

Because enhancers are understood to be orientation-independent as well as location-independent (Banerji et al., 1981), directional cloning was not undertaken for these ECRs. To obtain the expanded population of ECR sequences all that was required to R.E. digest them from pCRII's NotI and SpeI restriction sites. A benefit of this was that the process of amplifying the putative enhancer region was simplified by not having to clone in suitable restriction enzyme sites.

Region	Group	Location	Number of ECRs at specified stringency				
			100bp/ 70%	350bp/ 70%	350bp/ 75%	350bp /79%	350bp/ 80%
Total region IGSF11 to end of g19/g11	(ALL) 1,2,3	chr1:84259128-85558083					
IGSF11 to g9 isoform	1	chr1:84259128-84587321	29	**4**	3		0
g9 to start of g11/g19	2	chr1:84587321-85272240	113	24	12	7	**6**
g19 isoform (refseq)	3	chr1:85271704-85558083	50	13	**6**		0

Table 16 Genomic extent of ECR groups (all groups, top row) and number of ECRs (in bold/white box, where initial selection made) at various stringencies. Group 1 ECRs selected at a stringency of 350bps/70% similarity, Group 2 at 350bps/80% similarity and Group 3 at 350bps/70% similarity.

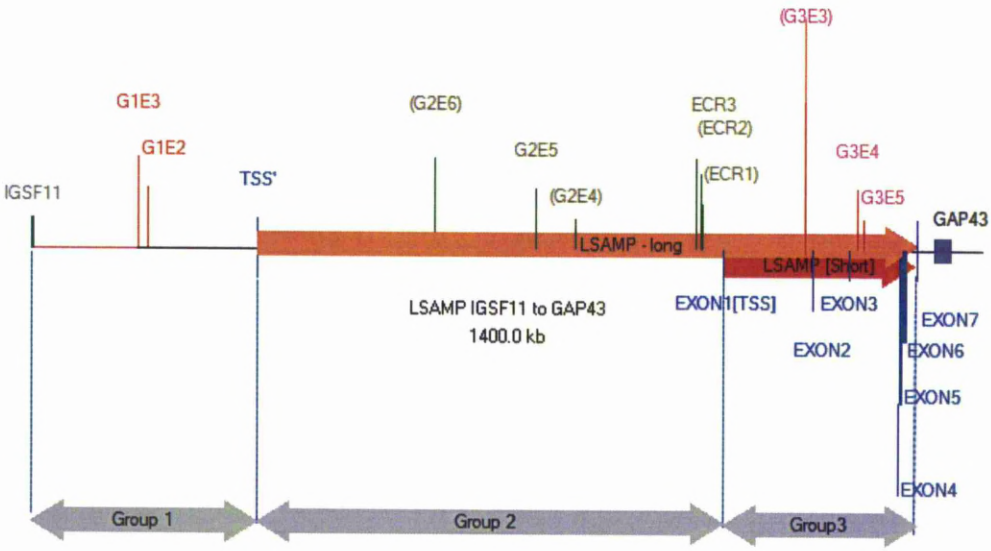


Figure 58 Isoforms of chicken *LSAMP*.

All isoforms of chicken *LSAMP* showing the location of the proximal transcription start site (EXON1 TSS) and upstream transcription start site (TSS'). Double-headed arrows indicate the genomic region Group1, Group2 and Group3. The location of the original ECRs (ECR1-ECR3, shown with old nomenclature) and the second tranche of 8 ECRs (GxEy). 'x' indicates the Group number (1, upstream of distal TSS; G2 between the distal and proximal TSSs; 3 within the transcription unit of the g11/g19 isoforms) and 'y' indicates the ECR number. ECRs resulting in no βgal expression are shown in round brackets. No GFP was obtained when GECR4 (G2E4) was used. Central line is 1.2 Mbp.

			ECR/Gene Name	Start	End	Length	% Identity
Upstream ECRs			G1E3	-84411211	84411576	365	76.2
				-84411576	84425789	14213	
			G1E2	-84425789	84426095	306	86.0
				-84426095	84587331	161236	
g9 (chLAMPb1) >> 84587331—85557493>>>	**TSS**		**g9 (chLAMPb1)**	**-84587331**	**85557493**	**970162**	
				-84587331	84846908	259577	
			G2E6	-84846908	84847366	458	82.1
				-84847366	84996713	149347	
			G2E5	-84996713	84997324	611	76.5
				-84997324	85055959	58635	
			G2E4	-85055959	85056467	508	76.6
				-85056467	85233806	177339	
	Original 3 ECRs		ECR3	-85233806	85234222	416	84.7
				-85234222	85239155	4933	
			ECR2	-85239155	85239299	144	83.7
				-85239299	85242299	3000	
			ECR1	-85242299	85242796	497	85.8
				-85242796	85272074	29278	
	g19 (chLAMPa1)	TSS'	**g19 (chLAMPa1)**	**-85271704**	**85558083**	**286379**	
		g11 (chLAMPa2)	**g11 (chLAMPa2)**	**-85272074**	**85557507**	**285433**	
				-85272074	85390785	118711	
			G3E3	-85390785	85391169	384	73.5
				-85391169	85469263	78094	
			G3E4	-85469263	85469580	317	83.3
				-85469580	85477683	8103	
			G3E5	-85477683	85478085	402	75.4

Table 17 ECRs initially selected over complete *LSAMP* test region with final selection of ECRs shown by name (GxEy). Location of ECRs in relation to the genome in chicken, length of ECR shown and percentage identity between chick (base organism) and human . Original three ECRs (ECR1,2 and 3) also shown.

5.4 RESULTS

5.4.1 Injections and Transfections into chick Neural Tube

The neural tube (n=132) and brain (n=32) of chick embryos at HH stages 10 through 14 were co-injected using the new *LSAMP* ECRs (GxEy) subsequently electroporated. The GFP expression that was observed in many of the embryos highlighted the regions of the embryos that were successfully injected and electroporated. Co-expression by the ECR-containing reporter plasmid was visualised as blue staining alongside the GFP expression and this was observed in the neural tube for 5 ECRs (G1E2, G1E3, G2E5, G3E4, G3E5) and included at least one ECR from each group. Blue staining was not observed in embryos injected with plasmids containing the 3ECRs (G2E4, G2E6, G3E3) (Table 16).

5.4.2 Sectioning the Chick Neural Tube to investigate location of βGal-derived blue stain.

To investigate if the individual ECRs were activated by different sub-populations of cells that could be identified by their location or migratory paths, cryostat sections of the blue-staining embryos were analysed. To assist orientation a diagram of the neural tube morphology is shown below in Figure 59.

ECR NAME (GX,ECRy)	ECR Length /bp	Orientation	Location	Inj Conc µg/µl	Hours Post Injection	Embryos Alive & Transfected	Embryos/ Active enhancer	% Transfected Embryos with enhancer
1,2	307	Fwd	NT	5.32	48	4	1	25
1,3	366	Rev	NT	2.3	48	14	8	57
2,1 (ECR1)	486	Fwd	NT	3.2	48	7	0	0
			NT		72	22	0	0
2,2 (ECR2)	205	Fwd	NT	4	48	3	0	0
2,3 (ECR3)	380	Fwd	NT	4.4	48	23	5	22
			NT		72	9	3	33
			NT		96	5	2	40
			Brain		72	3	0	0
2,4	509	Rev	NT	5.05	48	2 (No GFP)	0	0
			NT		72	6 (No GFP)	0	0
			Brain		72	2 (No GFP)	0	0
2,5	612	Rev	NT	3.9	48	7	1	14
			NT		72	8	3	37
2,6	459	Fwd	NT	3.0	72	6	0	0
3,3	385	Fwd	NT	5.0	72	7	0	0
			Brain		72	4	0	0
3,4	318	Fwd	NT	2.2	48	7	4	57
			Brain		72	3	0	0
3,5	403	Fwd	NT	2.88	72	8	6	75

Table 18 Summary of data for injections of the 8 ECRs into neural tube (NT) and brain showing ECR NAME (e.g. G1E2), length of ECR, injection concentration of p1229-based transgenic reporter vector, time incubated post injection, embryos alive and successfully transfected, embryos alive and transfected successfully at time of harvesting (48hrs or 72 hrs post injection), embryos containing active enhancer and so staining blue, percentage of embryos successfully transfected that contained an active enhancer. Original 3 ECRs included for clarity.

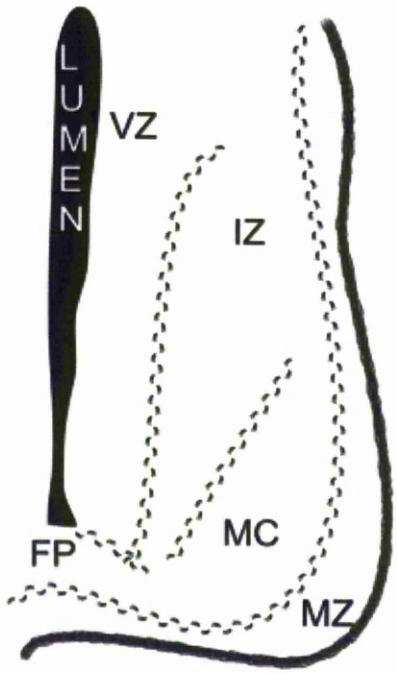


Figure 59 HH Stage 25 chick neural tube showing lumen of neural tube, ventral zone (VZ), intermediate zone (IZ), Medial Zone (MZ), Motor Column (MC) and Floor plate (FP).

5.4.3 Co-expression of GFP and LacZ/βGal derived Blue Staining

Blue staining in the NT of the embryos was driven by at least one ECR from each of the three groups. G1E2, G1E3, G2E3 (ECR3) from previous chapter, G2E5, G3E4 and G3E5 all produced blue staining (Table 16).

The whole-mount torso of the embryo injected into the central canal of the with G1E3 in Figure 60 is representative of the blue-stained embryos and shows the regions with the GFP expression Figure 60A highlighting the areas injected and electroporated and the corresponding blue staining in Figure 60B indicates the regions the reporter vector produced ECR-driven *LacZ*.

5.4.4 LSAMP ECR G1E3

Of the 16 embryos injected with G1E3, 88% (n=14) were expressed GFP and of these 57% (n=8) also exhibited blue staining due to βgal expression. A representative whole-mount embryo, in dorsal view, is shown in Figure 60. In order to analyse the cellular distribution across the neural tube 10 micron sections were made with differing expression patterns being observed.

In the section shown in Figure 61 GFP is expressed as a streak of highly localised staining about halfway down the dorso-ventral axis and running in a lateral direction. There are some fainter streaks more ventral than and dorsal this and there is a more diffuse arc of blue the dorsal end which could be neural crest cells. The amount of GFP staining evident is far greater than the amount of blue indicating either that the G1E3 enhancer becomes activated only in the stained subset of cells or, what is more probable, the *LSAMP* is not fully expressed at this stage, HH 23, of development.

The next section, shown in Figure 62, shows a fainter, less definite, streaked configuration that ranges from the dorsal (surface area) to as deep within the tissue as the GFP fluorescence is evident. An even more diffuse expression pattern is also apparent (Figure 63) although streaking is also discernible.

In all cases the blue staining is localised to the neural tube although GFP staining occurs in other areas such as the commissural neurons and the DRG which is especially evident in Figure 63. The DRG cells come from transfected neural crest cells that delaminate and migrate to form the DRGs. This indicates that these areas have been transfected but the G1E3 ECR is not being activated as an enhancer in this tissue at this time.

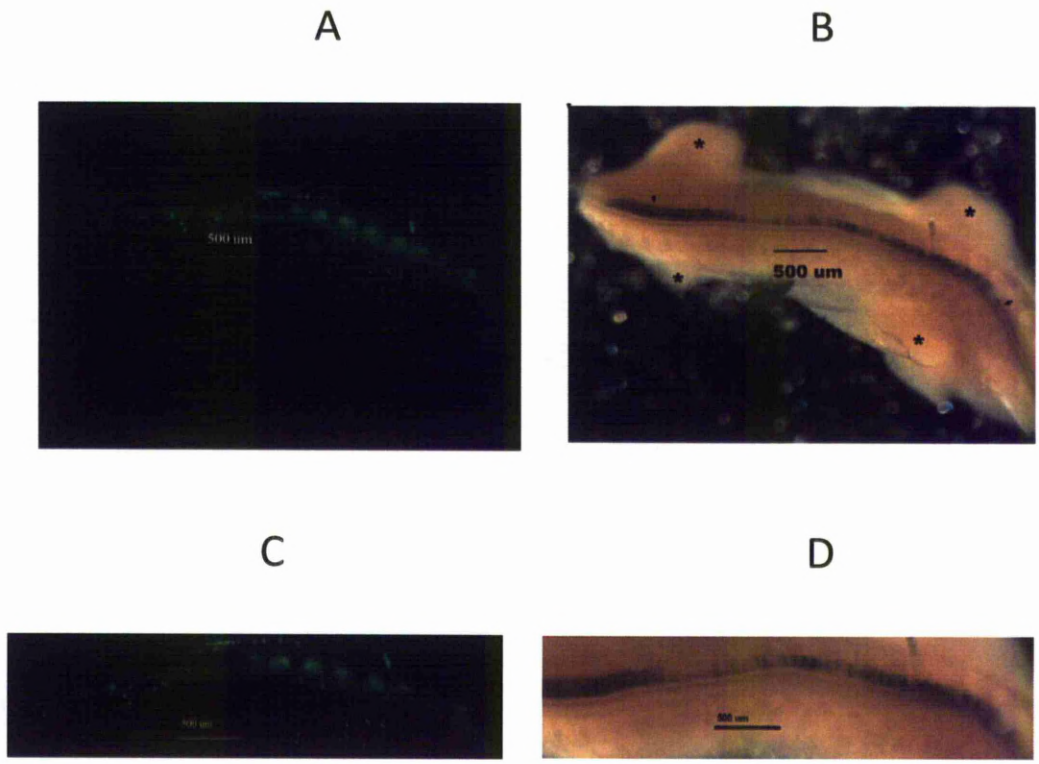


Figure 60 G1E3 HH stage 23 chick injected into Neural Tube at HH stage 14 with GFP plasmid and G1E3-containing p1229 reporter vector ,electroporated and incubated for 48 hours. (A,B) Overall view of whole-mount embryo (caudal, left ; rostral , right; limb buds identified with asterisks in(B)). (C,D) Rotated and cropped images of A and B respectively with the extent of the NT region in C and D is indicated by arrow heads in B. (A,C) Successful injection and electroporation visualised by green fluorescence. (B,D) blue chromogenic staining indicates that the ECR G1E3 functions as an enhancer.

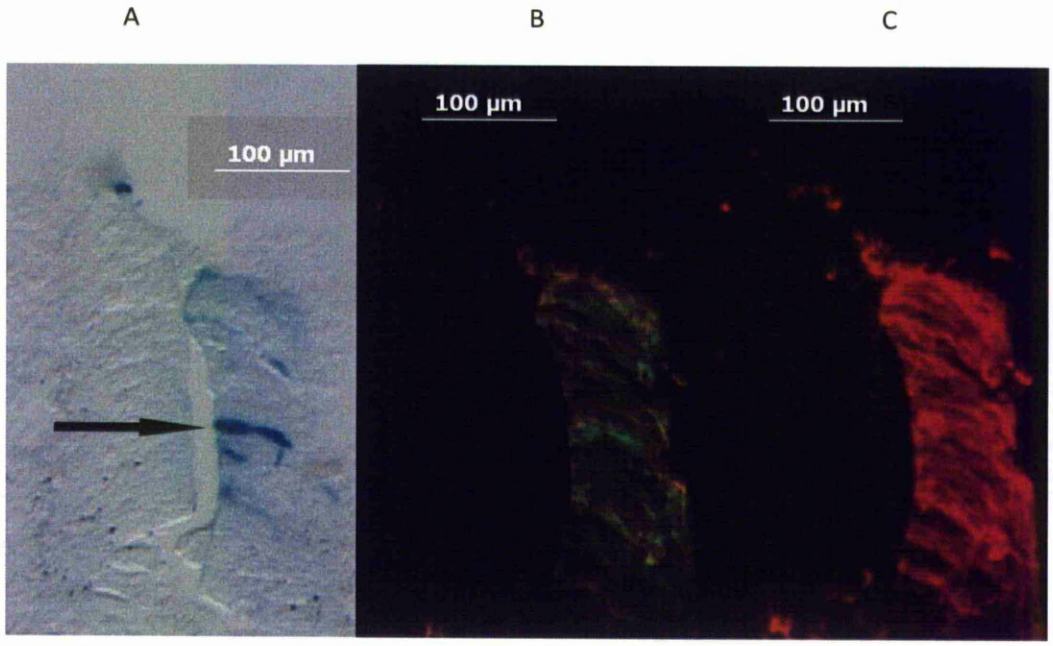


Figure 61 G1E3 10 micron section of E4 (HH stage 23) chick neural tube co-injected and electroporated at HH stage 14 chick with GFP plasmid and G1E3-containing p1229 plasmid construct, electroporated and incubated for 48 hours showing (A) intense staining (arrowed) but (B) with GFP intensity at similar levels to regions of faint, or even no visible blue indicating the specificity of the ECR.(C) Anti-GFP staining matches the GFP staining, and in this case the red fluorescent staining is less streaky than the green in image B.

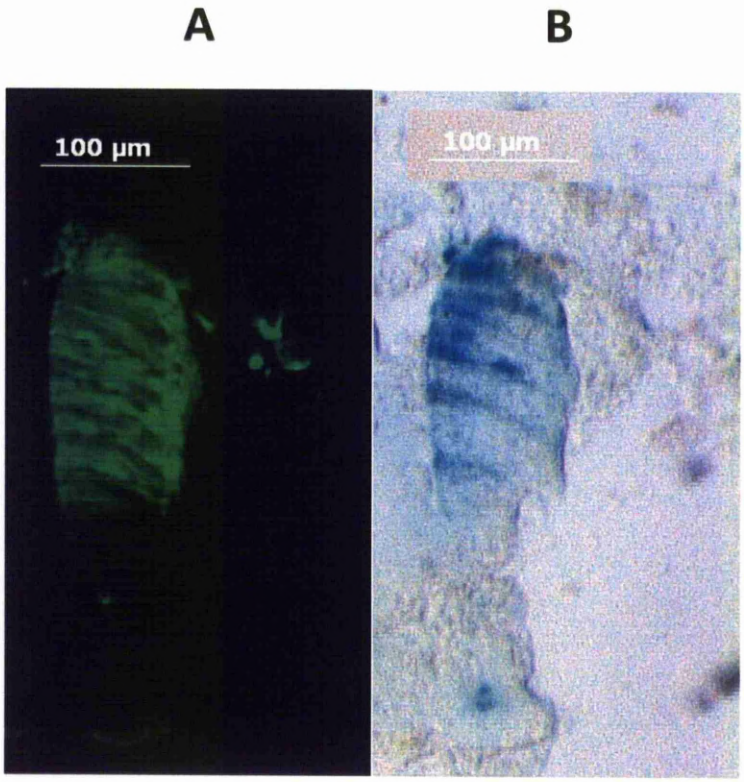


Figure 62 10 micron section of E4 (HH stage 24) chick neural tube co-injected and electroporated at HH stage 14 chick with GFP plasmid and G1E3-containing plasmid, electroporated and incubated for 48 hours. (A) Successful injection and electroporation visualised by green fluorescence. (B) Blue chromogenic staining evident in dorsal region of NT indicates that the ECR functioned as an enhancer at this stage. The streak-like staining pattern may indicate that these are radial glia.

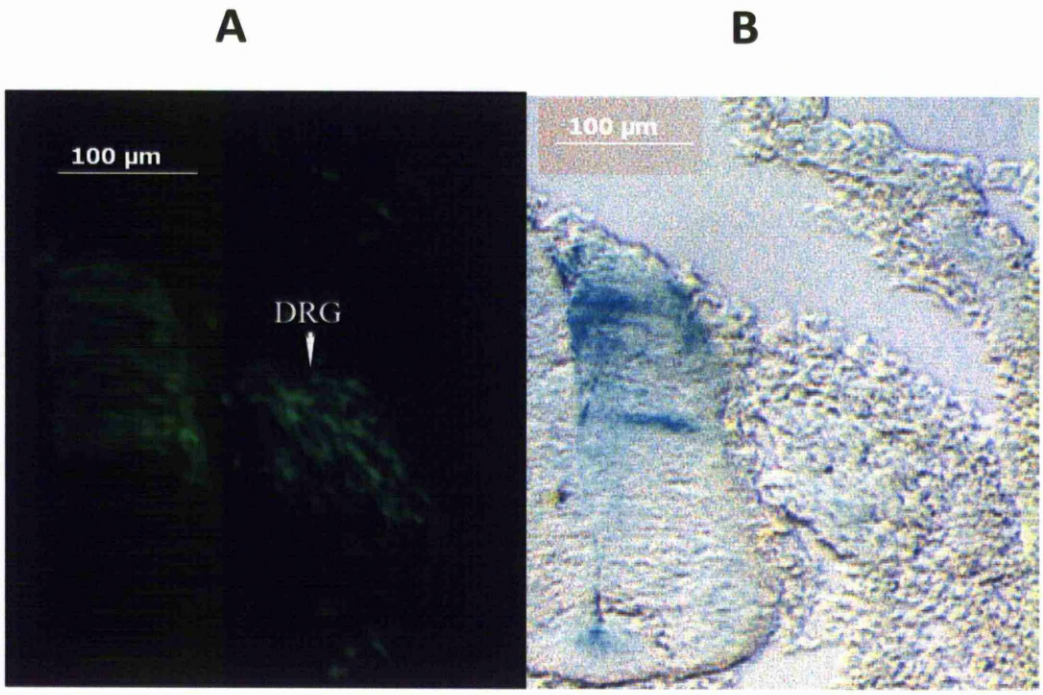


Figure 63 G1E3 Emb4R Sl8 C. Streaks of GFP partially correspond to diffuse blue staining. The dorsal root ganglion (DRG) is well labelled with GFP but with no sign of blue thus indicating the specificity of the ECR.

5.4.5 LSAMP ECR G2E5

With G2E5, 88% (n=15) of the 17 embryos injected expressed GFP, however only 27% (n=4) also expressed βgal driven blue staining.

Electroporation was successful as evidenced by the extensive GFP staining in the whole mount (Figure 64 A) but the blue staining appears fainter than would be expected from the example of G1E3 (Figure 63). This may be because G2E5 is activated in a different, and perhaps more discrete, subset of cells. The blue staining can be seen to be very intense in section and visible in dorsal medial and lateral regions (Figure 65 A,D), although GFP is visible throughout the hemi-section (Figure 65 C, E) indicating the success of the electroporation with the specificity of the blue staining indicating the preciseness of the G2E5 activity. GFP stained cell are also visible in marginal zone (Figure 65 B, E), where their orientation appears to be dorsal-ventral and there is corresponding blue staining within this region indicating that G2E5 is activated as an enhancer (Figure 65 D,G). The boxed *LacZ*-expressing region enlarged from Figure 65B shows the most dorsal blue staining (closed arrow) and the most intense blue (open arrows) (Figure 65 D) and GFP staining (Figure 65 E) and bright-field image (Figure 65 D) and GFP fluorescent (Figure 65 E) are shown as merged (Figure 65 F). The merged image (Figure 65 F) shows the co-localisation of the GFP fluorescence and the chromogenic blue staining. The strongest blue and GFP signals appear to be co-localised and no quenching of the GFP by the blue is observed (Figure 65 D,E,F) whereas the blue-fluorescent DAPI signal (Figure 65 G) appears quenched or masked indicating that the blue staining could be within the cell bodies.

In a separate embryo, as shown as a transverse plane piece of trunk in Figure 66, although the whole hemi-section of the neural tube and the DRG have been transfected as evidenced by the GFP only a single very strong blue band was generated in the dorsal region (arrowed in Figure 66 A). This again may show the early stage of *LSAMP* expression. In this case the

blue staining appears to be quenching or masking the GFP (arrowed in Figure 66 B) perhaps indicating that the blue staining is nearer the surface than the green fluorescence.

Figure 64 G2E5 images of HH stage 23 chick injected into Neural Tube at HH stage 14 with GFP plasmid and G2E5-containing p1229 transgenic plasmid construct and electroporated and incubated for 48 hours. (A,) Successful injection and electroporation visualised by green fluorescence. (B,) blue chromogenic staining indicates that the ECR functions as an enhancer.

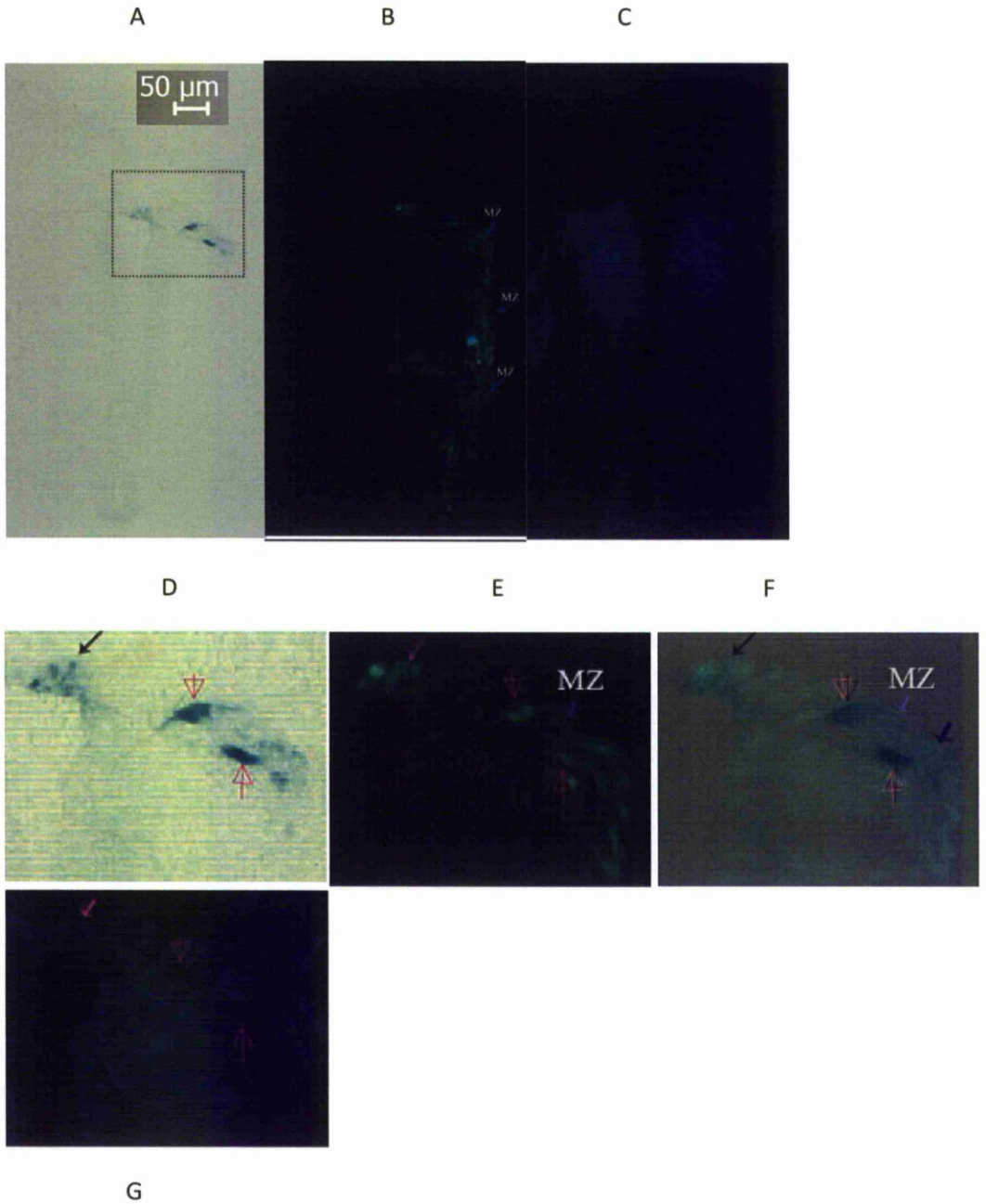


Figure 65 G2E5 Emb C T2 Sl4 C. Blue staining visible in dorsal region only (A,D) although GFP is visible throughout hemi-section (B E). GFP stained cell visible in marginal zone (arrowed as MZ) (B E), orientation appears to be dorsal-ventral and the corresponding blue staining can be seen (D and G). Boxed *LacZ*-expressing region enlarged from above (B) showing the most intense blue (D) (Black and open arrows) and GFP staining (E) co-localised . This co-localisation is shown in Image F, a merging of images D and E; the GFP and chromogenic blue staining can be seen in the MZ as indicated by pale blue and dark blue arrows. The DAPI signal (G) appears quenched or masked (pink open arrows) by the blue chromogenic stain.

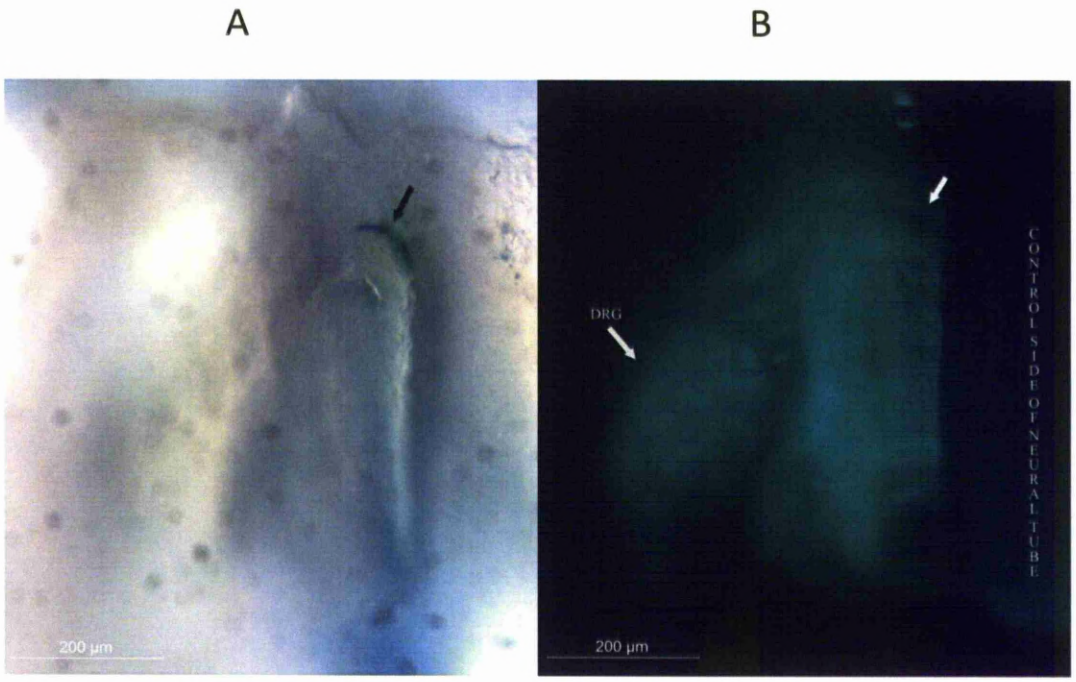


Figure 66 (A & B) Cross section view of piece of trunk showing discrete blue staining (arrowed). (B) The DRG is marked as is the electroporated "control side of neural tube".

5.4.6 LSAMP ECR G3E5

In some of the early sections GFP expression appeared to be poor and also what seemed to be the blue stain quenching the GFP. In an attempt to enhance the detection of the GFP signal, the sections were treated with GFP antibody and a secondary red-fluorescent stain. The results were disappointing in this instance with the original GFP signal being as strong, if not stronger, than the GFP antibody results (Figure 68 A (GFP), B (GFP antibody).

The blue staining in the whole mount image (Figure 67 B) is clumped and very intense. This intensity is reflected in section Figure 68 C but also probably by the dorsal location of the stained section. The blue staining is very localised although the GFP staining (Figure 68 A) and GFP antibody staining (Figure 68 B) indicate that the whole depth of the section was

successfully transfected, therefore the lack of blue staining was not due to restricted transfection.

To see if the staining patterns are co-localised overlays of the section images (Adobe Photoshop CS6) were made. The green, GFP image and the red, GFP antibody-stained images, Figure 68 A and B, were automatically aligned using Adobe Photoshop CS6 Auto-Align facility to produce Figure 69 C. The images containing the blue staining had to be manually aligned due to the blue image and the GFP and GFP-antibody images being less than the minimum 40% overlap required.

The overlay image shows that although there is extensive co-localisation of the GFP fluorescence and the red fluorescent GFP antibody indicator this co-localisation is not uniform (Figure 69 C) as indicated by the green (GFP fluorescence only), red (GFP antibody only) and yellow (co-localisation of GFP fluorescence and GFP antibody).

In a separate embryo, also positive for blue staining, the expression pattern was observed ventrally in both the ventricular zone (VZ) the intermediate zone (IZ), and medial to the motor column (MC) (Figure 70). The expression pattern here contrasts with the previous embryo where the expression was dorsal. This difference in the enhancer expression pattern between the two embryos shows that this enhancer probably drives expression from the dorsal to ventral regions of the NT.

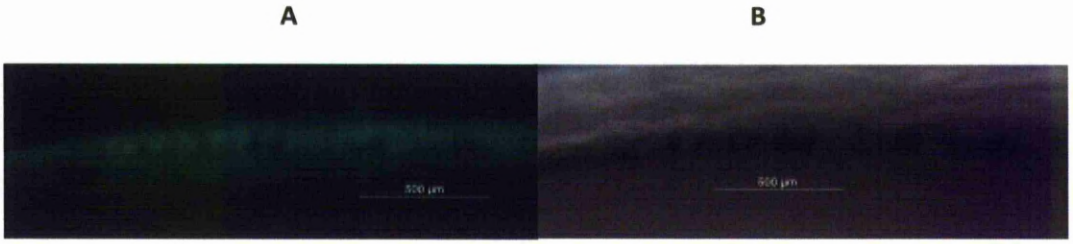


Figure 67 G3E5 images of HH stage 24 chick injected into Neural Tube at HH stage 14 with GFP plasmid (A,B) E/G3E5p1229:*LSAMP* transgenic plasmid construct and electroporated and incubated for 48 hours. (A) Successful injection and electroporation visualised by green fluorescence. (B,) blue chromogenic staining indicates that the ECRs functions as an enhancer.

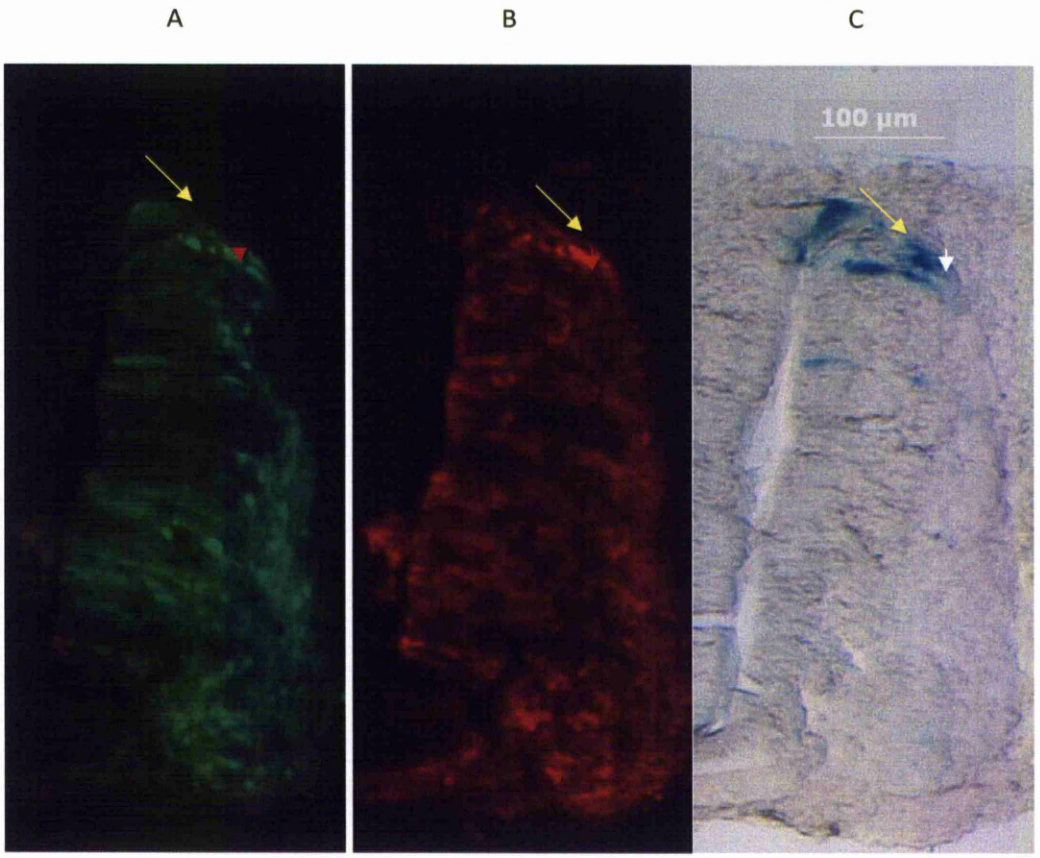


Figure 68 (G3E5 Emb B Sl A1 A) Comparison of 10 micron section of HH Stage 24 (E4) chick neural tube stained using three methods. (A) electroporation success indicator indicated by green fluorescent staining (B) GFP antibody staining indicating location of the GFP (C) blue chromogenic staining indicates that the G3E5 ECRs functions as an enhancer.

(A,B) GFP visible throughout Ventral zone, intermediate zone and motor column but blue staining mainly confined to dorsal region (yellow arrows). Note too that the intensity of blue staining (C) does not correlate to the intensity of the GFP staining (red or white arrow head (A,B)) .

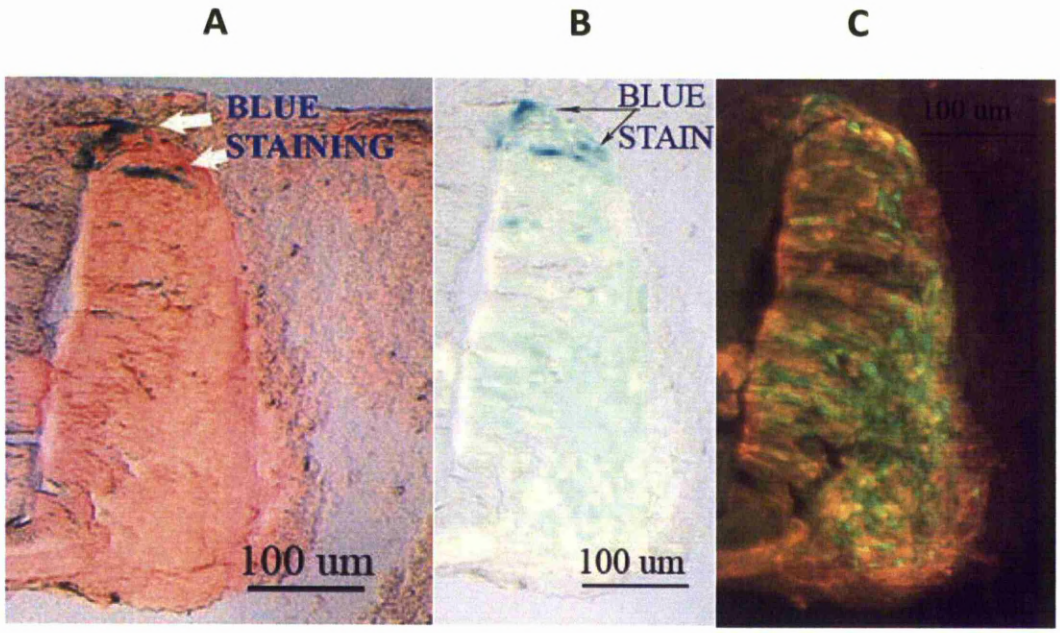


Figure 69 (G3E5 Emb B Sl A1 A) Comparison by overlay of 10 micron section of HH Stage 23 (E4) chick neural tube stained using three methods. (A) Electroporation success indicator indicated by GFP antibody stain indicated by red flurescence and blue chromogenic staining indicates that the G3E5 ECRs functions as an enhancer.

(B) GFP staining indicating location of the GFP and same blue chromogenic staining indicating that the G3E5 ECRs functions as an enhancer. (C) GFP (green fluorescence) and GFP antibody (red fluorescence) images overlaid to produce yellow colour where GFP fluorescence and red (GFP antibody indication) fluorescence coincide.

(A,B,C) GFP visible throughout Ventral zone, intermediate zone and motor column but blue staining (A,B) mainly confined to dorsal region (arrows). Note too that the intensity of blue staining is very localised.

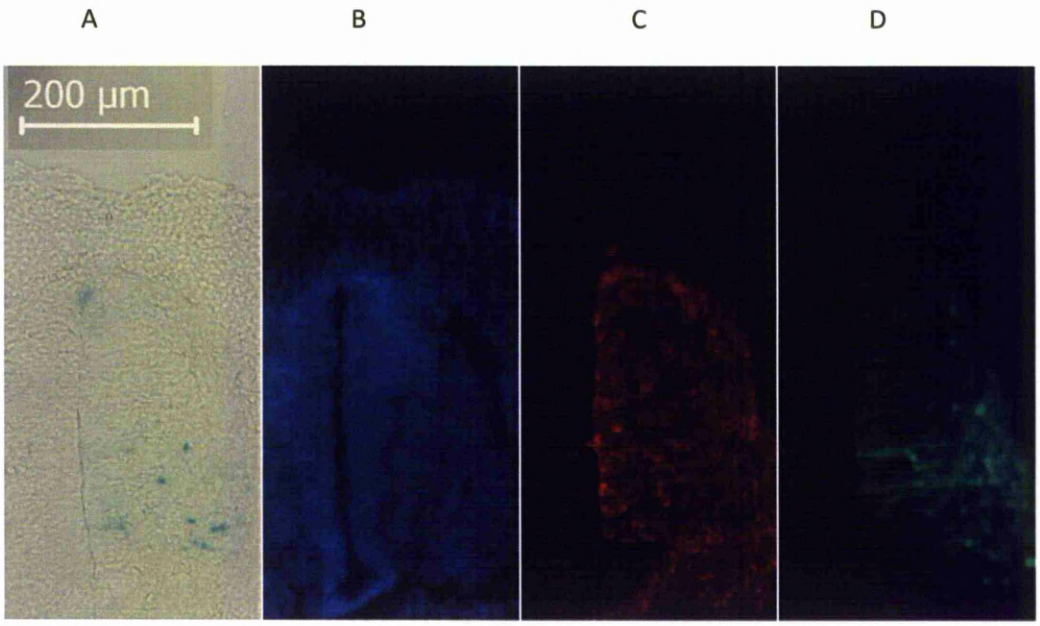


Figure 70 G3E5 Dots of blue staining are visible ventrally in the VZ and the IZ (A). Subsequent images show nuclei stained with DAPI (B), anti-GFP Ab with Alexa 2° Ab (C) and GFP fluorescence from the transfecting GFP (D).

5.4.7 LSAMP ECRs G1E2 and G3E4

Both the ECRs G1E2 (Figure 71) and G3E4 (Figure 72) were shown to be active in neural tube tissue at this stage of development with G1E2 appearing to be more prolific than G3E4. This may, once again, indicate differential activation of the ECRs in either or both a temporal or location-dependent manner.

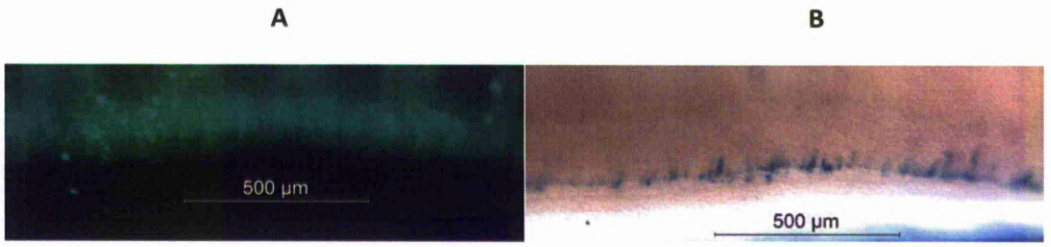


Figure 71 G1E2 Images of HH stage 14 chick injected into Neural Tube with GFP plasmid and (A,B) E/G1E2p1229:*LSAMP* transgenic plasmid construct and electroporated and incubated for 48 hours. (A,) Successful injection and electroporation visualised by green fluorescence. (B,) blue chromogenic staining indicates that the ECR functions as an enhancer.

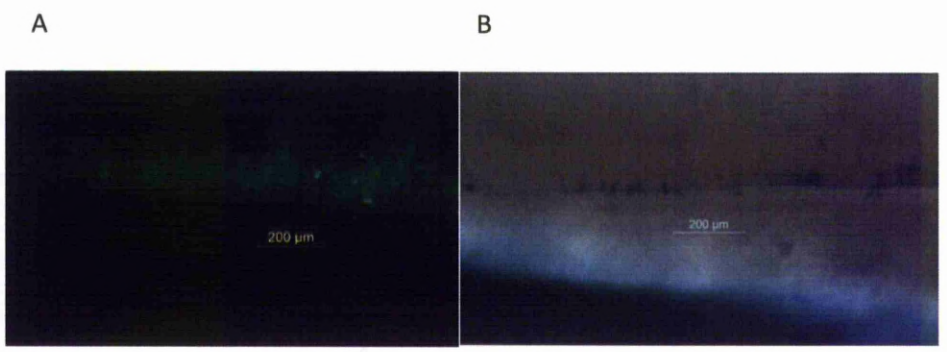


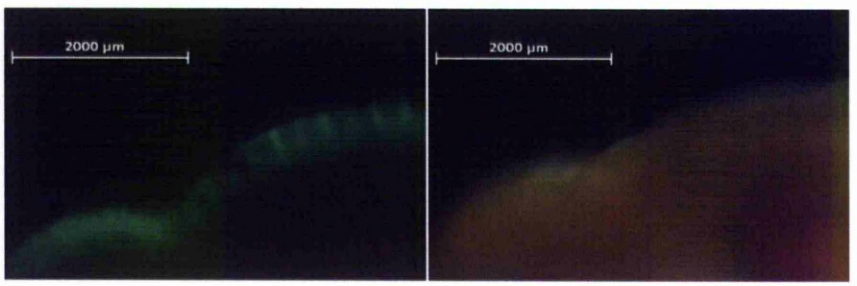
Figure 72 G3E4 Images of HH stage 14 chick injected into Neural Tube with GFP plasmid and (A,B) E/G3E4p1229:*LSAMP* transgenic plasmid construct and electroporated and incubated for 48 hours. (A,) Successful injection and electroporation visualised by green fluorescence. (B,) blue chromogenic staining indicates that the ECR functions as an enhancer.

5.5 *LSAMP* ECRs with no evidence of activity

No blue staining was observed from the following ECRs; G2E1, G2E6 and G3E3, although GFP expression was present.

5.5.1 LSAMP ECR G2E1

An example, G2E1, is shown in Figure 73 where it can be seen that although the GFP expression is strong no corresponding blue stain is observed.



A B

Figure 73 G2E1 HH stage 14 chick co- injected into Neural Tube with GFP plasmid and G2E1-*LSAMP* containing p1229 transgenic plasmid construct and electroporated and incubated for 48 hours. (A) Successful injection and electroporation visualised by green fluorescence. (B) lack of blue chromogenic staining indicates that the ECR did not function as an enhancer at this stage.

5.5.2 No GFP or LacZ-driven Blue Staining produced by LSAMP ECR G2E4

For one ECR, G2E4, no GFP or blue staining was observed. The G2E4 ECR appeared to quench the expression from the GFP plasmid although 23 embryos were injected into the NT and 6 were injected into the brain with plasmid containing this ECR. The reason for this is unknown, although it may have been due to a bad plasmid-prep as due to time constraints only one of these was run.

5.5.3 *Amount of Blue Staining increased with embryo's age*

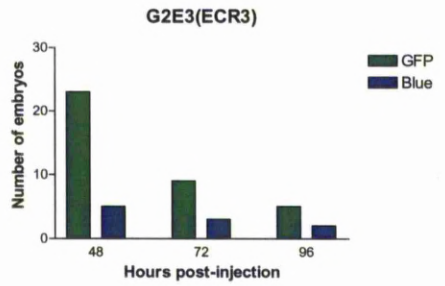
Embryos that expressed GFP were initially incubated for 48 hours (n=34) prior to staining but as viability improved and in order to see if staining continued and might increase in line with development stage the post-injection incubation time was extended to 72 hours (n=35) for two ECRs, because *LSAMP* expression is known to increase as the embryo develops. The percentage of the embryos that stained blue varied between enhancers, G3E5 (n=8) achieved 75% whereas G1E2 (n=4) only achieved 25% (Table 16) perhaps indicating the different activation pattern of the two ECRs.

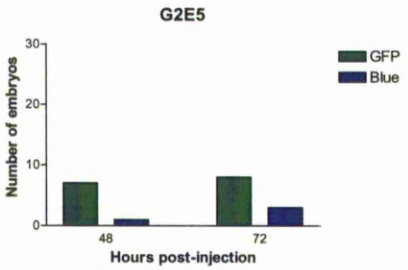
5.5.4 *Ratio of Blue-stained embryos to GFP-stained embryos increases with post-transfection incubation time.*

Although the numbers of embryos the were blue- and GFP- stained is a factor of the numbers processed and harvested at a particular time-point (e.g. 48 hours incubation), with the neural tube injections the proportion of embryos displaying blue staining increased with incubation time which is in-line with the expected increase in *LSAMP* expression over time at later embryonic stages (Figure 74).

As most of the embryos from the ECRs investigated in this chapter were all harvested at the same development stage ECR3-*LSAMP* (G2E3) from the previous chapter was used for comparison with G2E5 as embryos for both of these were harvested at, at least, two different development stages (Figure 74 C). The increase in blue was also found when all the blue-stained ECRs tested were analysed (Figure 74 D).

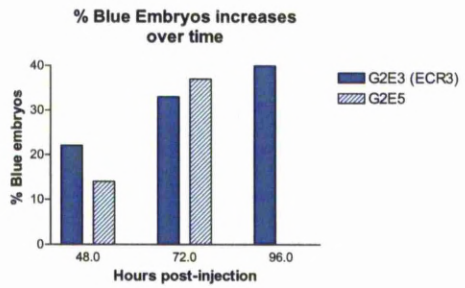
(A) (B)





(C)

(D)



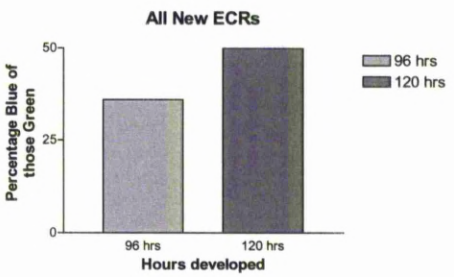


Figure 74 Percentage of embryos that stain blue increase with incubation time. (A,B) The absolute number of GFP-stained embryos that also stained blue depending on the number of embryos processed at that time. (C) The percentage of GFP positive embryos that are blue increases. (D) This also holds for all the ECRs tested in this chapter.

5.5.5 Injection and transfections into Brain by a subset of the new LSAMP ECRs failed to produce any LacZ/βGal derived blue staining.

Applying the same rationale and methodology as described in the previous chapter injections were carried out into the brain using the following ECRs; G2E4, G3E3, G3E4.

Once again the GFP-staining in the brain was found to be extensive indicating that the procedure functioned efficiently and that exploration of enhancer activity in the brain would be practical using this system. Even where GFP staining was strong no βGal-derived blue staining was perceived in the brain as described in previous chapter.

Although 26 injections and electroporations were made in this chapter into the brain using, G3E3 and G3E4, the survival and success rate was poor with only 7 embryos in total being GFP positive and surviving to harvest. This does not include G2E4 which was problematic and did not express GFP even in the NT. No expression of blue was observed in any of embryos (Table 17).

ECR	H.H. Stage Injection	No. Injected	Hours incubated post-injection	No. Surviving	% surviving	H.H. Stage Harvest	GFP	βGal	% blue of those GFP+ve
G2E4*	13/14	6	72	4	66	25	0	0	0
G3E3	12	13	72	7	58	25	4	0	0
G3E4	14	13	72	4	31	25	3	0	0
Total		32		**15**	**47**		**7**	**0**	0
Total – G2E4 omitted		26		**11**	**42**		**7**	**0**	0

Table 19 Summary of data for injection into brain at various HH stages showing number of embryos injected that expressed GFP. None of these expressed blue staining. *G2E4 did not express GFP.

5.5.6 *Transcription Factor Binding Site located within the new LSAMP ECRs.*

A large number of TFBS were found to be conserved in the ECRs and are shown in Table 24 in the Appendix. The most interesting perhaps are the TFBSs that appear in the ECRs that did not drive expression. These are marked by salmon-pink rows and the negative ECRs are shown in red columns. The following group of ECRs, LEF1_Q2, LEF1_Q2_01 and LEF1B_01 appeared in G2E4, G2E6 and also in G2E1 (ECR1-*LSAMP*). Lymphoid enhancer-binding factor 1 appears to be involved in the negative control of Interleukin-4 LEF-1 (Hebenstreit et al., 2008) but it is also active in positive control (van Genderen et al., 1994).Further investigation into this group of TFBSs and the LEF1 transcription factor may prove productive.

5.5.7 *Analysis of Blue Staining by Heatmap Diagram of transfected hemi-section of chick Neural Tube.*

An analysis of the regions of blue-stained cells was undertaken. The hemi-section of the NT was divided into 8 sections (4 medial and 4 lateral) and scored for any occurrence of blue (Figure 75). The results showed that the most dorsal of the medial sectors contained blue staining most frequently and that the number of stained sectors reduced in a lateral and ventral manner (Table 18, Figure 76)

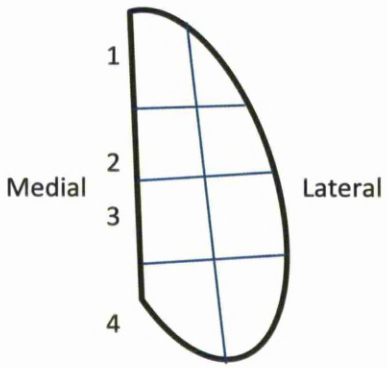


Figure 75 Diagram of NT hemi-section as used to enumerate sections containing LacZ-driven blue. If a section contained any blue it was marked as positive.

	G1E3 n=18 (%)		G2E3 n=18 (%)		G2E5 n=21 (%)		G3E5 n=25 (%)		TOTAL n=82 (%)	
Col/Row	M	L	M	L	M	L	M	L	M	L
1	100	66.7	88.9	61.1	95.2	42.9	92	24	93.9	46.3
2	55.6	55.6	33.3	83.3	0	0	0	0	19.5	30.5
3	50	27.8	50	72.2	0	0	20	8	28	24.4
4	11.1	16.7	44.4	22.2	0	0	4	4	13.4	9.8

Table 20 Percentage of GFP-stained sections that expressed blue staining (M = medial, L = lateral) All sections that were analysed were GFP positive throughout.

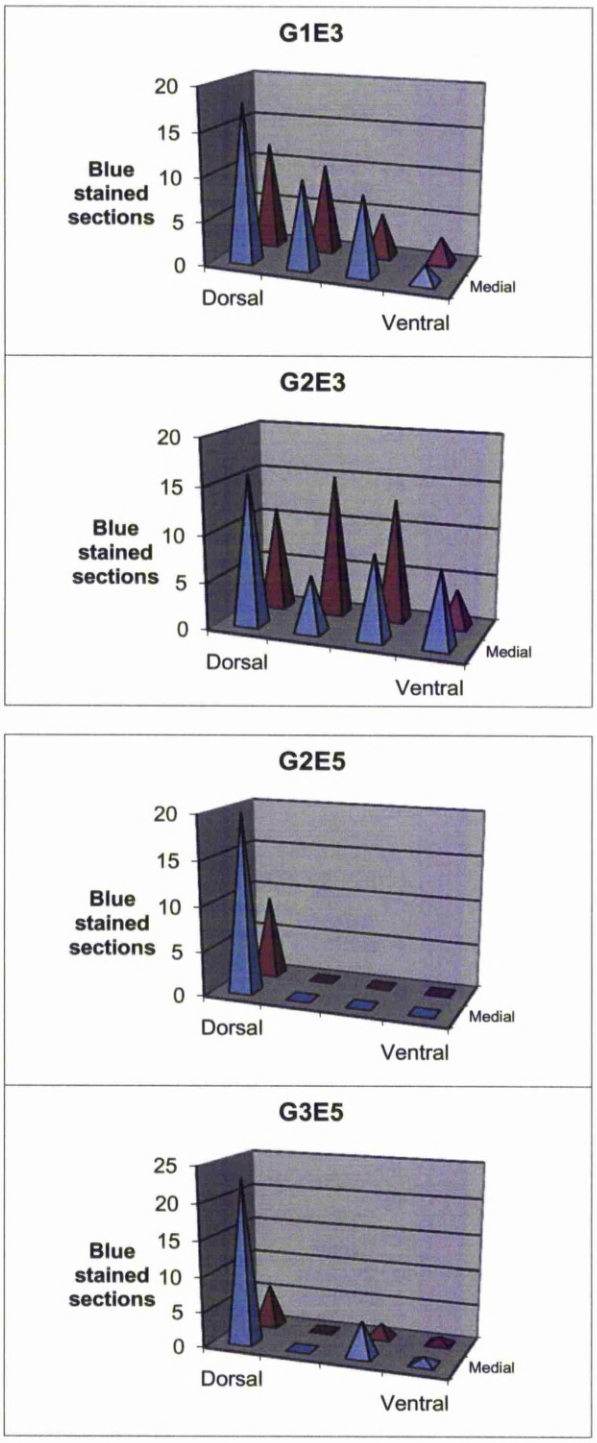


Figure 76 Heatmap of blue staining in the NT. The hemi-section of the NT was divided into 8 sections (4 medial and 4 lateral). If a sector included any blue stain it was counted in.

In a small number of cases (G2E3 (ECR3)) and G3E5) blue staining was observed ventrally in both the ventricular zone (VZ) the intermediate zone (IZ), and medial to the motor column (MC) Figure 70. These cells may be migrating neurons.

5.6 Transcription Factor classes in all ECRs

Within the ECRs tested in this thesis for enhancer activity 211 different conserved TFBSs were identified (Table 27). These 211 TFBSs can be grouped into about 31 classes (Table 28) . The class for one of these TFBSs, COMP1 was not identified and there are three TFBS that refer to PolyA sequences although it is moot if they are TFBS as they are not in the TRANSFAC Professional database (Matys et al., 2006) , however they are included in these data.

Of the 211 different TFBS 55% (n=117) were found only in the ECRs that drove *LacZ*/ βGal expression, 31% (n=67) were found only in the ECRs that did not drive such expression and 13% (27) were found in both categories (Figure 77).

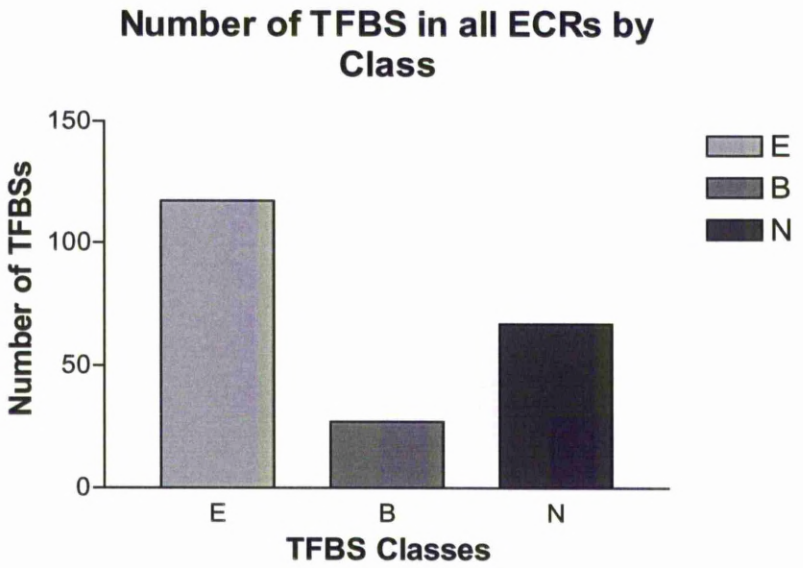


Figure 77 Number of the 211 discrete TFBS, shown by class, in all *LSAMP*-related ECRs tested in this thesis.

Key: E = TFBSs located in ECRs shown to have only enhancer activity

B = TFBSs located in ECRs that had both enhancer and no enhancer activity in this tissue and at this stage of development

N = TFBSs located in ECRs that only showed no enhancer activity in this tissue and at this stage of development

The HOX class of TFBSs was the most highly represented at 25% (n=53) of the total TFBSs but these were spread fairly evenly between the different categories of ECR; 9.5% (n=20) were in ECRs that only drove *LacZ*/βGal expression, 8.5% (n=18) were found only in the ECRs that did not drive such expression and 7.1% (n=15) were found in both categories. The BZIP TFBS class came in a close second with 9% (n=19) being in ECRs that only drove *LacZ*/βGal expression, 3.3% (n=7) were found only in the ECRs that did not drive such expression and 1.4% (n=3) were found in both categories, but were less evenly distributed between the categories.

Of the 33 TFBS classes 27% (n=9) were found only in the ECRs that drove *LacZ*/ βGal expression. Of these two classes stood out; the E2F class at 36% (n=12) contained the

greatest number of individual TFBSs with the HISTONE class being second at 21% (n=7) . Conversely 18% (n=6) of the TFBS classes were found in only in ECRs that did not drive expression. The HMG class contained the highest number of TFBSs 1.9% (n=4) with the PolyA group containing 1.4% (n=3) of all TFBs.

5.7 DISCUSSION

Following generation of the reporter plasmids, injection and electroporation was performed as previously described with injection in chick embryo neural tube and also brain at HH stage 10 to HH stage 14. Five of the eight ECRs produced blue staining following injections into the NT as had ECR3 (G2E3) (Table 16). Of the 32 injections made using the new tranche of ECRs into the brain, a low number of 8 (25%) were GFP positive however no βgal-derived blue staining was produced. This was not entirely unexpected because 5 of the GFP positive injections were co-injected with a G3E3-containing p1229 which had failed to produced blue in the NT, and so may be a repressor and therefore only 3 of the GFP positive injections contained an ECR, G3E4, that resulted in βgal derived blue staining in the neural tube.

The ECRs that functioned as enhancers in the neural tube varied in robustness, with rates ranging from 14% to 75% of the transfected embryos expressing βgal. Most of the embryos were harvested following 48 hours incubation except for G2E5 where harvesting occurred at both 48 hours (Stage 23) (n=7) and 72 hours (Stage 25) (n=8) post-incubation. G2E3 (ECR3) GFP positive embryos had been harvested at 48 hrs (n=23), 72 hrs (n=9) and 96 hrs (n=5) post incubation.

Of the embryos that were GFP positive for G2E3(ECR3) at 48 hours post-injection(n=23) and 72 hours post-injection(n=9) and for G2E5 (48 hrs PI ,n=7 and 72 hrs PI n=8) , βgal-expressing embryos increasing from 20% of those that were also GFP-stained (n= 6 of 30) at 48hrs (HH stage 23) to 35% (n=6 of 17) at 72 hour HH Stage 25) indicating that this ECR

may become more active in line with increased incubation time as had been found with ECR3. This increase in activity in the neural tube is in line with Kimura's findings that at later stages of development expression of LAMP was seen at all levels the spinal cord (Kimura et al., 2001). The exact development stage is not specified in Kimura's report , from the context of the paper it seems that this is sometime after HH stage 25 which is at a more advanced stage than most of the staining described in this project occurred.

5.7.1 ECR Orientation or length did not affect enhancer activation

Enhancers are orientation independent (Banerji et al., 1981) and this was borne out by the fact that the forward orientation G1E2,G2E3,G3E4,G3E5 and the reverse orientation G1E3 and G2E5 all drove *LacZ* and so produced blue embryos. Ideally some enhancers would have been tested in both orientations to establish this beyond doubt.

The length of the ECR had no specific effect on the activation of the reporter vector (Table 16). Both the shortest ECR, the 307 bp G1E2, and the longest ECR, the 612 bp G2E5, activated the reporter vector. Bracketed by these lengths were eight other ECRs, four of which activated the reporter vector and four that did not.

The combination of ECR length and the ECR orientation also appears to have no effect because both G1E2 which is 307 bp and in forward orientation and G1E3 which is 366 bp and in reverse orientation both activate the reporter vector, and at the other extreme, G2E5 at 612 bp in reverse orientation, as has already been noted, does activate the reporter vector whereas G2E6, the second longest ECR, apart from the questionable G2E4, is 318 bp and does not cause activation of the enhancer (Table 16).

5.7.2 *Certain ECR classes may indicate greater likelihood of enhancer activity.*

In the ECRs that were found only in the group that drove *LacZ*/βGal expression the TFBS for both the E2F (n=12) and the HISTONE (n=7) TF classes contained the greatest number of individual TFBSs . On the contrary the HMG class contained the highest number of TFBSs (n=4) in the group that was shown not to drive reporter activity.

When investigating future ECRs for enhancer activity these data would suggest that the presence of TFBSs for the E2F and HISTONE classes of TFs should be a positive selection criterion whereas the presence of the HMG TF class should be a negative criterion.

5.7.3 *Lower levels of Plasmid concentration may be effective*

Injection concentration of the reporter vector had been recommended at up to 6 μg/μl per plasmid, but this was not always achieved. The highest concentration of 5.32 μg/μl (G1E2) and the lowest concentration of 2.2 μg/μl (G3E4) both activated the reporter vector so it is probable that this lower value could be used for all experiments and being less viscous would be easier to inject. Dilution tests could be run to determine the lowest concentration of plasmid that would produce consistent results.

5.7.4 *GFP may have been quenched by one ECR.*

For some unknown reason the GFP expression from the GFP-expression plasmid appeared to be quenched by the p1229 reporter vector that contained the G2E4 ECR. This GFP is expressed under the ubiquitous β-actin promoter and so it appeared unusual that it should be quenched. Could the p1229 plasmid be absorbing all the transcription factors required by β-actin core promoter in the GFP-expressing vector? Probably not. Another possibilty is that some toxin was introduced by the G2E4:p1229 construct. A more mundane possibility is that the transfection failed for this ECR although as 29 embryos were injected, in two

separate experiments, using two different mixes of plasmid and GFP-expressing plasmid this would appear to be unlikely. Another possibility is that although two plasmid preparations were run for the p1229 plasmid containing the G2E4 insert, only one produced any output and a new overnight expansion had to be run. It is possible that there was a problem with the second plasmid preparation although at the time there was nothing to indicate this. The sequencing of the p1229 plasmid with the G2E4 insert produced the expected output. If this was to be pursued further, new endotoxin-free plasmid prep should be made of G2E4:p1229. If this quenches the GFP then further investigation should take place, for example, co-injecting the G2E4:p1229 plasmid with another p1229/GFP mix that is known to function and observe if quenching still occurs. In this case an investigation into the TFBSs in the G2E4 and the GFP-expressing plasmid

5.7.5 Expression pattern is initially higher in Dorsal region but becomes more extensive with age

The analysis of the regions of blue-stained cells showed that the most dorsal of the medial sectors contained blue staining most frequently and that the number of stained sectors reduced in a lateral and ventral manner (Table 18, Figure 76). If the cells being electroporated originate from the ventral zone (VZ) and migrate laterally into the intermediate zone (IZ) this medial-lateral gradient would alter in a temporal manner as the cells migrate their numbers would increase laterally and decrease medially. In a small number of cases (G2E3 (ECR3)) and G3E5) blue staining was observed ventrally in both the ventricular zone (VZ) the intermediate zone (IZ), and medial to the motor column (MC) Figure 70. These cells may be migrating neurons and axons.

G1E3 showed a gradual reduction in expression both laterally and ventrally compared with G2E5. G2E5 was confined to the dorsal region whereas G3E5 had high expression dorso-

medially; however expression dorso-laterally was less than would have been expected from the other two examples G1E3 and G2E3. G3E5 also exhibited some expression ventrally.

This variance in expression pattern could be as a result of the enhancers being activated in differing sub-populations of cells or it could be as a result of inconsistent transfection or even of greater transfection success in the ventral region of the G2E5-injected embryos. Another possibility for this variation is that it could be as a result of different maturation stages of various cells and that had the staining been allowed to carry on past HH stage 25 an increasing proportion of the cells may have expressed βgal or as a result more success .

5.7.6 Transfected Region

Of note was the difference in the extent of the GFP expression in the neural tube between the Hb9 proof-of-concept transfection as against the GFP expression for the ECR transfections. When the injection and electroporation was carried out using the Hb9 enhancer (Chapter 3) the GFP signal was strongly visible throughout the complete hemi-section and within the floor-plate (FP, c.f. diagram Figure 59), however when the same protocol was followed for the *LSAMP* ECRs the GFP signal in some cases did not appear to reach full ventral extent and in no instance was the floor plate GFP-stained. This led to the concern that some of the cells that could expresses β-Gal were not transfected.

On closer inspection there did not appear to be a full correlation between the intensity, or in some cases, the presence of GFP and the blue staining. In Figure 68(A,B) the GFP is visible throughout the hemi-section. GFP expression is faint in the dorsal region where the βGal expression is strongest and is very strong ventrally where the βgal expression is virtually non-existent. The GFP antibody staining (red) shows a slightly different pattern of expression than the green GFP but this may be a result of GFP at a deeper level than where the antibody has bound indicating that the antibody is more sensitive.

In section, most of the neural tubes were found to have blue-stained cells at the roof plate/neural tube junction (Figure 62). G2E5 was shown to drive expression here too (Figure 65) and G3E5 has an especially large expanse in this region (Figure 68 C).

5.7.7 No blue observed following injections into Brain

This low survival and success rate is possibly an operator-effect due learning how to inject into and electroporate the brain (Table 17).

Although *LSAMP* is expressed in the anterior mid-brain at HH stage 14 (Kimura et al., 2001) it may be that these particular enhancers (G3E3 (n=4), G3E4 (n=3)) may not be activated in these cells at this time, or that the *LSAMP* expression window was too brief to trigger the βGal expression in sufficient quantities to be detectable. Another possibility is that although the GFP staining was intense, indicating that transfection had been successful, due to the large lumen the concentration of the enhancer was insufficient to result in observable blue staining. Whether these enhancers are active or not in the brain has not been proven due to the small number of transfected embryos achieved.

It must also be noted that there was a very high mortality rate associated with these injections into the brain with only a low number 8 of 32 injected, i.e. 25%, surviving and expressing GFP as compared to the high number of 65 of 132 i.e. 49% achieved with neural tube injections. While it is understandable that piercing the embryonic brain tissue with a relatively large hypodermic needle would present an enormous insult to this highly complex organ and so lead to death, the total lack of success in achieving blue staining may be a product of the bent rod electrodes failing to concentrate the electric charge sufficiently to focus the effect of the transfection to provide sufficient β-gal expression. Of course there is always the possibility that none of the putative enhancers were activated by the electroporated cells at this particular stage in development.

Some of the ECRs were shown to drive expression in cells of the chick embryonic neural tube. The LacZ driven blue staining varied between ECRs but also between embryos injected with the same ECRs. Migration appeared to result in blue-staining cells being found more laterally as a function of time. So far no correlation between TFBS and the ECRs that drove LacZ expression has been established

Chapter 6

6 General Discussion

One of the major ways employed by eukaryotic cells to control gene expression is by the action of the regulatory sequences known as enhancers and repressors. An understanding of the spatial and temporal activation of these sequences would help our understanding of gene control. Furthermore polymorphisms in these control regions have been associated with disease and a way to rapidly identify if these polymorphisms were affecting gene expression would be a useful diagnostic tool. In order to attempt to rapidly identify when and where enhancers are being activated a chick-based model was proposed. Using the ECR Browser, putative enhancer sequences were identified by cross-species comparison of the genomes to expose evolutionary conserved regions. By cloning an ECR into a suitable reporter vector and transfecting cells *in-vivo,* the activation of the enhancer by the cell can be detected. Following a successful proof-of-concept run using the *Hb9* gene, *LSAMP* was selected as the first gene to investigate because of its role in development and its implication in mental health disorders and cancer. A number of ECRs were examined and six of these were shown to be enhancers.

6.1 Advantages of this chick-based model and the positive results achieved.

This method has been demonstrated to function successfully by using ECRs located near the *LSAMP* gene to identify activated enhancers. These enhancers were shown to be active in the same region as LAMP is normally found, indicating that these enhancers are probably involved with *LSAMP* expression.

This is a cheap and relatively rapid method with results being available following several hours of embryos staining, in which little operator involvement is required.

The GEISHA (Gallus Expression in-situ analysis) database is being built by means of whole mount in-situ hybridisation using chick embryos (Bell et al., 2004) . The protocols they use detect mRNA and microRNA is quite rapid, and they have built a database showing the expression pattern of many genes, but not *LSAMP*. The genes have been categorised into gene families and what could be of use in the future is their data on the temporal and spatial expression of transcription factors.

The VISTA enhancer database (Visel et al., 2007) experimentally validate human and mouse ECRs in transgenic mice and as of early May 2012 they had tested 1729 sequences, 891 of which exhibited enhancer activity. Similar to the method used in this work the ECRs are selected on cross-species conservation and they also use putative enhancer markers. As the VISTA system uses mice that are harvested at E11.5 the murine-based system is intrinsically slower than the chick-based system as the equivalent chick age is ~E3.5 (Schneider and Norton, 1979) . It may appear that because chick is more distant evolutionary from human as chick/human they had the last common ancestor at ~300MYA and human/mouse was ~100 MYA. Due to the rate of genomic divergence for chick being estimated at 0.16 per million years and that of mouse being 0.9 per million years it is concluded that the organisation of the human genome is closer to the chick than the mouse (Burt et al., 1999). In the ECR Browser, by comparing chick with human gives the advantage that due to the evolutionary distance and yet stasis between the two species, primitive control regions are identified. As these regions are conserved between the species, using this tool to investigate within these regions is likely to prove most useful.

6.2 IgLONs and *LSAMP* continue to be implicated in disorders and disease

Data revealing the relationship between IgLON polymorphisms and disease continue to be discovered with polymorphisms in CEPU1/NTM having been implicated by association with lower intelligence and with late-onset Alzheimer's disease (Pan et al., 2011). And various polymorphisms in *NEGR1*, the gene that codes for Kilon, having been associated with obesity (Elks et al., 2010, Renstrom et al., 2009, Bauer et al., 2009) and a specific polymorphism (rs2815752) in *NEGR1* was associated with higher waist circumference, but also, unexpectedly, with lower body mass index (Haupt et al., 2010).

New evidence, has emerged, indicating that a 2.1 Mb region in 3q13.31 that contains *LSAMP* was frequently deleted in osteosarcomas (Kresse et al., 2009). This indicates that *LSAMP* may have a role as a potential tumour suppressor gene in osteosarcomas.

Further evidence supporting *LSAMP*'s role as a potential tumour suppressor gene in osteosarcomas was provided by the fact that a deletion localized to a region within the *LSAMP* was identified, along with mutations in other regions, as being present in 6 out of 11 osteosarcomas investigated (Yen et al., 2009). It was found that the deletion of *LSAMP* promoted proliferation of normal osteoblasts by regulation of apoptotic and cell-cycle transcripts and also VEGF receptor 1 (Pasic et al., 2010).

Additionally, a SNP-array analysis on 300 diagnostic and 41 adult and pediatric leukemia samples, identified, along with other aberrations, a recurrent minimal deletion region at 3q13.31, a expanse containing only *LSAMP* (Kuhn et al., 2012).

A novel functional relationship between LAMP, glucocorticoid receptors, and cognitive functioning has been identified (Qiu et al., 2010). This was achieved by comparing *wt* mice and *Lsamp*$^{(-/-)}$ mice for spatial memory performance, hippocampal synaptic plasticity, and

stress-related responses. The *Lsamp*$^{(-/-)}$ mice exhibited a pronounced decrease in spatial memory acquisition and reduced hippocampal CA1 region long-term potentiation over the *wt* controls. Reduced expression of mineralocorticoid receptor (MR) transcripts in the hippocampus and reduction in the corticosterone-induced, MR-mediated non-genomic modulatory effects on CA1 synaptic transmission were also observed.

Knock-out mouse models are in agreement with earlier findings (Catania et al., 2008) that show lower anxiety in *Lsamp*-deficient mice and furthermore show a decrease in agonistic behaviour (Innos et al., 2011). *Lsamp*$^{(-/-)}$mice were also found to be less able to adapt to novel stressful environments and interact in a socially normal way (Innos et al., 2012).

Polymorphisms in *LSAMP* have also been implicated in major depressive disorder and panic disorder in humans (Koido et al., 2012). The *LSAMP* genomic region of 591 patients suffering from major depressive disorder or panic disorder was compared against 384 healthy volunteers and 21 SNPs were identified. None of these SNPs were located in any of the ECRs investigated in this thesis. Only one of these SNPs, rs9830559, was found to be in an ECR (Location, Human genome build 19 ; chr3:115595803-115596470 Chick genome build 3 ; chr1:85514662-85515326) when chick was compared against human at stringency of 100bp/70% identity.

6.3 Limitations and Caveats

Because enhancers and repressors sequences need not be located near the gene they are moderating there is no guarantee that these sequences are affecting the nearest gene. Although the expression pattern of the blue staining matched that expected of *LSAMP* this is still circumstantial and is not a complete guarantee.

There are many different breeds of chicken and although the white Leghorn is used most often in reproductive studies (Bahr, 2008) and was used during this thesis it is not identical

to UCD 001 whose genome has been, incompletely, sequenced (Lee et al., 2003). The release 2.1 (May 2006) of the chicken genome was throughout this work. The origin of the whole-genomic chick DNA used experimentally is unknown. These potential differences between the documented genome, the chick DNA from which the ECRs were copied and the actual breed of chick embryos used leads to the possibility of dissimilarities, the extent of which is unidentified.

An even greater source of difference is that chickens are evidently not humans. We diverged from our common ancestor about 310 million years ago and although many of the basic systems will be similar there will be many differences. This means that anything discovered in the chick model, as with all models, will only be an indication of what happens in other organisms. The greater the difference in evolutionary time between the model and organism under investigation, typically, the greater the variances will be.

This method, so far, has used cross-species comparison to find evolutionary conserved regions under the assumption that they are putative enhancers. However, this would thus miss enhancers that are human-specific or are rapidly evolving which by their nature are not conserved. Using input from a predictive bioinformatics methodology such as (Arunachalam et al., 2010, Fernández and Miranda-Saavedra, 2012) could be used to predict enhancer sequences and our system could then be used to test the bioinformatics prediction.

The *LSAMP* gene, although important and interesting, proved a challenging candidate to start this investigation due to its large size and complexity, and a smaller simpler gene would probably have been more appropriate although positive results were obtained from *LSAMP*.

So far expression of *LSAMP* has not been detected in the brain using this method, but as the GFP fluorescence indicates transfection is effective this shows that there is no fundamental defect in the transfection process.

6.4 Enhancements made to the methodology during the course of the work

Knowledge gained from early experiments and availability of new equipment allowed the methodology to be improved. Due to the high percentage yolks breaking when transferring the fertilized eggs into the hammock, "windowing" of the eggs replaced "hammocking" . To reduce the risk of egg shell breakage when removing standard Sellotape, Magic Scotch Tape was used instead. Both of these changes reduced the amount of loss due breakage.

When finance became available a mains powered electroporator was purchased to replace the old, battery powered one. This had more facilities and gave greater control to the operator. An active-humidity incubator was also purchased to replace the passive-humidity incubator allowing for the incubation process to approach nearer to the natural conditions.

6.5 Proposed Future Experiments and Directions

The blue, βgal driven chromagenic staining although widely used in this context, possibly due to its permanency has the major disadvantages that it is not visible before the staining process and that the staining process is lethal to the embryos. If the p1229 plasmid was altered to express a fluorescent marker instead of LacZ the expression could be detected while the embryos were still alive allowing the decision at what stage to dissect the embryo to be better informed. Because there would be no stain-development phase the processing time would also be reduced thus allowing results to be obtained more rapidly.

It is unknown if some of the ECRs cloned into the reporter vector were repressors or instead were enhancers that were activated at too low a level to drive the *LacZ* sufficiently to result in staining, or simply aren't regulatory sequences. One way to increase expression is to clone multiple copies of the enhancer sequence into the reporter vector. In one series of experiments up to four identical copies of an enhancer sequence has been shown to

increase expression, beyond that, expression is adversely affected (Kumar et al., 1986). Nakano used three copies of the enhancer sequences to increase reporter gene expression (Nakano et al., 2005).

Some enhancers work in co-operation to increase expression and some will only function in the presence of another (Maekawa et al., 1989). In order to test if the enhancers act co-operatively to stimulate promoter activity and to ascertain if some of the ECRs that did not drive reporter expression when alone would function in the presence of another, I would propose that multiple ECRs be cloned in combination, in a step-wise manner, into the reporter vector. An increase in reporter expression would indicate that co-operation between the ECRs has occurred whereas decreased or complete quenching of reporter expression could indicate that most recent ECR added has a repressor function.

The amount of literature on TFs is growing and the FANTOM consortium have gathered existing and *de-novo* data into the FANTOM4 EDGE database (Ravasi et al., 2010). The data on TFs is collected from the following sources; Evolutionary conserved TFBS from sources that include JASPAR and TRANSFAC, results from miRNAs used to target transcript predictions from the EIMMO prediction server (Wang and El Naqa, 2008), published protein: DNA interactions, published TF:TF interactions from various sources such as the Database of Interacting Proteins -DIP (Xenarios et al., 2000) , Biomolecular Interaction Network Database- BIND (Bader and Hogue, 2000)and Human Protein Reference Database – HPRD (Prasad et al., 2009), in-house and published ChIP-chip data and published ChIP-seq data and in-house siRNA knock-out and miRNA over-expression on THP-1 cell-line. Filtering is available to produce, for example, a list of TFs of which there is experimental support. Using this database could be used to help interpret the TFs and TFBS bindings that moderate the activation of the enhancers/repressors.

Interpretation of the current systems results are qualitative and very subjective as it does not quantify the results which are subject to multiple variables such as the success of the electroporation, the subset of cells transfected, the developmental stage of the embryo, the developmental rate. In order to use this system effectively it would be important to quantify the level at which the reporter gene is driven otherwise much information on the co-operativity of enhancers and the effect of repressors could not easily be captured. There are various assays, such as western blotting that can be used to quantify protein expression but these are not rapid systems. A quicker method is to use a static and variable fluorescent assay such as a dual-luciferase reporter (Alcaraz-Perez et al., 2008) . In this system a control luciferase, of one colour, can be driven by a constitutive promoter and another luciferase, of different colour, can be driven by a promoter under control of the ECRs under test. By normalising the signals from experiments using differing combinations of ECRs on the control luciferase the effect of the various enhancers/repressors could be quantified (Ray and Gambhir, 2007, Fortes and Razquin, 2009). Another method would entail replacing the constitutive promoter with a promoter under the control of a single enhancer or a different combination of enhancers to those controlling the other luciferase's promoter.

The promoter sequence in the p1229 reporter vector is a human β-globin minimal promoter, and although it has been demonstrated in this work that it functions, the use of plasmids containing the reporter gene under the control of the native promoter sequence should provide a more intrinsic response to the putative enhancer sequence. In this way the generated signal would be more quantitative because the native promoter should respond to the enhancer/repressor sequences innately. It would also be expected that in most cases the reporter expression would be driven at a higher rate under control of the native promoter.

Although it is generally understood that enhancers and repressors function in a direction-independent manner, it would be worthwhile to prove this for the ECRs that have already been analysed. This would entail inserting the ECRs into the reporter vector in reverse orientation to their current direction and retesting. If any of the enhancers no longer give a signal or the ECRs that previously did not give a signal do so now, further investigation would be required to ascertain if they are direction-dependant. If this double-orientation testing was to become the norm it might be desirable to re-investigating the feasibility of using the directional, or non-directional "Gateway" method, to allow easier reporter vector generation.

6.5.1 Proposal to reduce infection in embryos

Although the fertilized hens' eggs used in the work described in the previous chapters, were produced by specialist producers who maintain a high level of cleanliness, eggs by their nature are far from aseptic. Some of the eggs we received had visible faecal contamination on their shells and even on apparently clean eggs there is, presumably a myriad of fungal and bacteria coating the shell. Although eggs possess protection they are not totally immune from external infection, especially by fungus, particularly if incubation is delayed. This contamination can also reduce egg viability considerably (Cook et al., 2003). During the work described in this thesis it was found that the viability of the eggs deteriorated with time and also that whole experimental runs were ruined by what appeared to be a fungal infection. As this infection tended to occur during the autumn it was assumed that the fungus originated on dying leaves, however, upon reflection, it may be contamination on the eggshell. Furthermore, as it will be desirable to allow the embryos develop for longer periods this problem will no doubt become more significant.

Upon receipt of a batch of eggs, our procedure is to place them, in their papier-mâché trays directly into a passively humidified fridge at 12°C. In order to potentially reduce the risk of

fungal infection and even possibly increase the viability of the stored eggs the procedure should be changed, at least for a trial period.

Under the proposed procedure, on receipt of the eggs, they should be washed in 70% ethanol, all evidence of faecal matter to be removed and they should be placed in the fridge in sterilised trays. The current washing in 70% ethanol prior to opening the egg should be continued. Following electroporation embryos should be covered in an antibiotic and fungicide solution in order to reduce infection (Croteau and Kania, 2011).

6.5.2 Expanding the tool's use into human ECRs

Whilst investigating how chick ECRs in chick moderate chick gene expression would be valuable in its own right one of the desirable uses for this tool would be to investigate human ECRs in the chick model. The first attempt at this would be just substituting the human equivalent ECRs for the chick ECRs shown to be enhancers and see if they still function as enhancers. By comparing the TFBSs in the two sequences extra data on the TFs that are involved in regulation may be obtained.

Following this it would be appropriate, by using site specific mutagenesis to investigate how mutations within the regulatory sequence affect expression. If this is shown to function, SNPs identified by GWAS surveys for diseases with specific genetic causes, such as the DeNT GWAS (Breen et al., 2011), should be used to establish if the SNPs can be shown to affect gene expression.

6.5.3 Transfection methods

Electroporation still appears to be an effective and preferred method transfecting chick cells and our general method i.e. electrode gap, voltage, pulse length and ~ one second interval is in line with other's methods used when electroporating the chick neural tube (Croteau and

Kania, 2011) although platinum or platinum/iridium wire electrodes seem to be more commonly used than the gold-plated electrodes used for all but the earliest experiments previously described. Using the solid wire seems to allow more flexibility in cleaning and maintaining the electrodes in that the tips of the negative electrodes can be cut off, say, every dozen electroporations and the positive electrode can be cleaned using a forceps to scrape off any build-up. The gold-plated electrodes have the advantage of solidity but they are very expensive (£350 per set), have an estimated life of ~1,500 pulse and have to be cleaned gently to prevent ablation of the gold plating.

Electroporation has several disadvantages the first of which is that not all the expected cells are electroporated leading to a mosaic transfection pattern. Why this happens is not fully understood but it is possible that the shape of the cell and its orientation to the electric field prevents the cell reaching the transmembrane potential that triggers electroporation (Pucihar et al., 2006). To fully elucidate the expression pattern resultant from a particular enhancer requires the results from a number of transfected embryos to generate a composite view. A second problem is that generally the expression by the transgenes is transitory as the plasmids typically are not be integrated into the genome resulting in a diluted signal as the plasmids are distributed between a proliferating cell population. This problem would become more acute as the embryos are allowed to develop for longer with its concommitant increase in cell divisions. Cell divisions occur in a temporal (Fujita et al., 1962, Wilson, 1973) and location dependant manner (Fujita, 1962, Smith and Schoenwolf, 1988) making exact calculations difficult, however an estimate (Table 19) would indicate that increase in cells for an injection made after 48 hours incubation and stained a further 48 hours later would be a power of 5.5 so a 100 cells would become 9,500 in that 48 hour period, resulting in a possible 95:1 dilution of the reporter vector.

Day	1	2	3	4	5	6
Cell-cycle/hours	5	7	8	9	13	15

Divisions/day	4.5	3	3	2.5	2	1.5

Table 21. Frequency of cell cycles from embryonic day 1 to day 6 in chick.

The combination of a provirus vector, such as Replication-Competent ASLV long terminal repeat (LTR) with a Splice acceptor (RCAS) and virus-competent eggs (Federspiel and Hughes, 1997)I facilitates the reporter's integration into the genome and the cross-infection of cells overcomes the short-term expression and mosaic problems. Therefore if loss of reporter vector expression diminished with increased development time an assessment of this system should be undertaken with a view to implementing it, if it appeared appropriate.

6.6 Overall conclusion

Evidence of *LSAMP's* position as a tumour suppression gene and its role in anxiety, and by implication in deviant social behaviour (Innos et al., 2011)is accumulating, as too is the indication of disease being related to polymorphisms associated with this gene.

This highlights the need for a diagnostic method to rapidly identify the effect of polymorphisms on this or any other gene.

This thesis has demonstrated that the ECR Browser can be used to identify chick ECRs that act as enhancers of gene expression and gives a starting point for further development.

APPENDIX

7.1 Reagents

Name/Function	Source	Catalogue or model number	Notes/comments
Agarose for Electrophoresis	Sigma	A5304	
Agarose for routine use	Sigma	A9539	
Ampicillin stock	100 µg/µl		
BlueJuice	Invitrogen	10816-015	10X gel loading buffer
BSA			Bovine serum albumin
Cloned AMV	Invitrogen	12328-032	
DAPI Fluorescent nucleic acid stain	Sigma	D8417	4'6-Diamidino-2-phenylindole dihydrochloride.
DNA ladder 1kb	Invitrogen	15628-019	
DNA ladder 1kb plus	Invitrogen	10787-018	
DNA, whole-chick, genomic	Merck-Novagen	69233-3	249 ng/µl
dNTP 10 mM	Invitrogen	12328-032	
Dulbecco's PBS	Sigma	D8662	
EDTA 0.5M	Gibco	155 75-038	
Eggs, Fertilised chicken's	Lea's Lane Farm The Pinfold Centre Farmfresh, Preston	White Leghorn	
Glycerol	BDH	44448	
HBSS	Sigma	H9394	
HBSS (-)	Gibco	14170	No Ca^{2+} No Mg^{-}
HBSS (+)	Gibco	24020	
Ink	Pelikan Fount Ink		Used at 7% in PBS
IPTG	Invitrogen	15529-019	
LB Broth	Sigma	L3022	
loading buffer	Invitrogen	TrackIt™ Cyan/Orange 6X 10482-028	10bp-1kbp
loading buffer	Invitrogen	BlueJuice 10X Cat. No.10816-015	
loading buffer	Invitrogen	10482-035 TrackIt™ Cyan/Yellow 6X	100bp-10kbp
MNR2	University of Iowa	81.5C10	Monoclonal Ab with a pan MNR2/HB9 immunohistology pattern
molecular-weight marker	Invitrogen	TrackIt™ 100 bp DNA Ladder 10488-058	100 – 1500 bp
molecular-weight marker	Invitrogen	TrackIt™ TrackIt™ 50 bp DNA Ladder 10488-043	50 to 800 bp
NBT (Nitro blue trizoleum)	Roche	G2451025	
NP40 IGEPAL	Sigma	I-3021 CA630	
OCT medium	Bright	Cryo-M-Bed medium	Mounting medium
pCRII vector	Invitrogen	45-0641	
Phase-lock gel Heavy	Eppendorf	0032 007.953	
Plasmid midi-prep	Qiagen	Plasmid midi-prep 12143	
Plasmid midi-prep	Sigma Aldrich	GenElute plasmid mini-prep PLN-70	Endotoxin-free
Plasmid mini-prep	Promega	Pure Yield mini-prep A1223	
Plasmid mini-prep	Sigma Aldrich	GenElute plasmid mini-prep PLN-70	
Potassium ferricyanide			
Potassium ferrocyanide			

S.O.C. medium	Invitrogen	46-0700	
Sodium deoxycholate			
SyBR Green I	Sigma	S-9430	
TOPO Cloning kit	Invitrogen	46-0800	
Wizard SV gel and PCR clean-up system	Promega	A9281	
β-D-galactopyranoside	Sigma	B4252	X-gal

Table 22 Reagents used with suppliers name and catalogue number.

7.2 Equipment

Name/Function	Source	Catalogue or model number	Notes/comments
Capillary glass tubing	Warner instruments	Mod G150TF-3	OD 1.5mm ID 1.17mm Length 7.5cm
Centrifuge	Sanyo	MSE Mistral 3000i	Swinging bucket
Centrifuge	Sanyo	MSE Micro Centaur	Bench
Centrifuge	Heraeus Instruments	Biofuge Stratus	High speed
Centrifuge	Heraeus Instruments	Biofuge Pico	Bench
Cloned AMV	Invitrogen	12328-032	
Electrodes, 3mm bent gold-plated	Harvard Apparatus, Inc.	BTX Model 514, #45-0116	
Electrodes, 5mm bent gold-plated	Harvard Apparatus, Inc.	BTX Model 512, #45-0115	
Electroporator	Ovodyne electroporator	TSS20	
Electroporator, battery-powered	Digitimer Ltd	Isolated stimulator Model DS2	
Hot block	Techne DriBlock	DB-2A	
Incubator - shaking	New Brunswick Scientific , Edison, N.J.	Controlled Environment Shaker Incubator Model G25	
Incubator – cell/tissue	Heraeus	BB6060 CU	
Micropipette puller	Sutter Instruments	Model P-87 Flaming/Brown	
Egg fridge (storage)	Fison		Temperature/Humidity 12°C /Passive Humidity
Incubator; Egg incubator (pre-injection)	Brinsea Products, Sandford, BS25 5RA	Multihatch MkII	Temperature/Humidity 38.5°C / Passive
Incubator; Cell	Heraeus	BB6060 CU	Temperature/Humidity 37°C/ Passive humidity CO_2 @ 5%
Incubator; Cell	Sanyo CO_2 Incubator	MCO-15AC	Temperature/Humidity 37°C/ Passive humidity CO_2 @ 5%
Incubator; Cell – used for eggs	R S Biotech	Galaxy Standard	Temperature/Humidity 37°C/ Passive humidity
Incubator; Egg and Humidity Management Module	Brinsea Products, Sandford, BS25 5RA	OvaEasy 380A H22	Temperature/Humidity 37.8 °C – 38.0 °C / RH 35%. Note: Rocking mechanism in incubator has been disabled.
Microscope	Leica	Leica Leitz DMRB	Fluorescent microscope Metamorph™ software. LAS software V2 & V3 Camera: DFC420C
Microscope - Inverted	Leica	Leica Leitz DMIRB	Fluorescent microscope. LAS software V3 Camera: DFC420C
Microscope - Dissecting	Leica	Leica Leitz LB M165FC	Fluorescent microscope. LAS software V2,V3,V4 Camera: DFC420C Transmitted Light Base External halogen goose-neck illuminator
Microscope – Dissecting	Motic	Motic Stereo Dissecting	Motic transilluminated

	base External halogen goose-neck illuminator

Table 23 Equipment used during work on thesis.

7.3 Solutions and Buffer Recipes

7.3.1 Millonig's Phosphate Buffer

Sodium Phosphate, monobasic $NaH_2PO_4.H_2O$ 12.68g/l

Sodium Phosphate, dibasic Salt NaH_2PO_4 43.80g/l

$NaH_2PO_4.H_2O$and NaH_2PO_4were dissolved in 900ml of de-ionized water (DIW) and pH adjusted with HCl or NaOH to pH7.4 at room temperature. Volume made up to 1 litre with DIW.

7.3.2 Millonig's Phosphate Buffer pH 7.4 (0.125M)

0.4M Phosphate Buffer diluted 3:7.

7.3.3 4% Paraformaldehyde

ddH_2O	35 ml
Paraformaldehyde	2g
NaOH	4-5 drops
Phosphate Buffer 0.4M	10ml

35 ml of distilled H_2O was added to 2g paraformaldehyde and stirred constantly whilst being heated gently to 65°C. 5M NaOH was added dropwise until solution cleared. When solution had cooled to room temperature the volume was brought up to 50ml using ~10ml 0.4M Phosphate Buffer. pH was assayed and adjusted to pH 7. 10 ml aliquots were stored at -20°C.

7.3.4 Luduena's Fix

100ml volume	
Glutaraldehyde	8% (v/v)
Sucrose	0.12 M
$CaCl_2$	0.5 mM
Phosphate buffer pH 7.4	75 mM

7.4 Bacteriological Media

7.4.1 Luria Broth

25g of Luria Broth base added per litre of DIW H_2O and autoclaved at 121°C for 15 minutes

7.4.2 Luria Agar

37g of LB Agar base added to 1 litre of DIW H_2O and autoclaved at 121°C for 15 minutes

7.5 Antibiotics

	Stock	Working concentration
Ampicillin	100 mg/ml	50 µg/ml
Kanamycin	30 mg/ml	50 µg/ml

Antibiotics were dissolved in DIW and sterilized through 0.2 µm filter and aliquoted and stored at -20°C.

7.6 Tris-HCL 1M

121.1g Tris base dissolved in 800 ml ddH_2O at room temperature and pH adjusted to required value. Volume topped up to 1 litre and sterilised by autoclaving 121 °C/15 minutes.

7.7 Electrophoresis Buffer for Agarose Gel Electrophoresis

Buffer was prepared as a 50X concentrate stock solution and stored at room temperature. For use it is diluted to 1X in de-ionized water.

7.7.1 50X Tris-Acetate (TAE) Buffer Stock.

Tris (TRIZMA base)	242g
0.5M EDTA pH8	8.0 ml
Glacial Acetic acid	57.1 ml
H_2O	make up to 1 litre.

7.7.2 Agarose-gel running buffer - 1X TAE

50X TAE	100 ml
H_2O	make up to 5 litres.

pH and adjusted to pH 8 if required.

7.8 Embryo Staining

7.8.1 Sodium Azide 10% stock solution

10g of sodium azide in 100 ml of distilled H_2O. this is stored at room temperature.

7.8.2 Embryo Staining Solution

Phosphate Buffer	0.1 mM pH7.3
$MgCl_2$	2mM
Sodium deoxycholate	1%
Potassium ferrocyanide	5mM
Potassium ferricyanide	5mM
NP-40	0.02%
X-gal	1mg/ml (added prior to use).

7.8.3 Detergent Rinse

Phosphate buffer	0.1M
$MgCl_2$	2mM
Sodium Deoxycholate	0.01%
NP-40	0.02%

7.8.4 Diluent solution

1% BSA,
0.12M Pi (Diluted Millonig's)
0.1% Triton,
0.1% NaN_3

Table 24 Solution mixes used in the work carried out for this thesis.

7.9 Primers used for generating ECR1, ECR2 and ECR3 in *LSAMP*

Primer name	Primer sequences as ordered	Total Primer length	Total Tm/°C	Primer3 Primer length	Primer3 Product Size/bp	Primer3 Tm/°C
up*LSAMP* ECR1 fwd	aaatGCGGCCGCaatctgaacgttgccctgtc	32	84.8	20	486	60
up*LSAMP* ECR1 rev	atctACTAGTagttcagtaaagctgcccaaa	31	66.4	21		59
up*LSAMP* ECR2 fwd	aggaGCGGCCGCagtccagctgccaacagaat Due to strong hairpin, changed from this: agctGCGGCCGCagtccagctgccaacagaat	32	85.8	20	205	60
up*LSAMP* ECR2 rev	ctgcACTAGTgcagaaggcctgaaggaaat Chenged due to very strong dimer: cacgACTAGTgcagaaggcctgaaggaaat	30	74.7	20		61
up*LSAMP* ECR3 fwd	ataaGCGGCCGCctgtctgcctgtctgtctgc	32	83.8	20	380	59.76
up*LSAMP* ECR3 rev	ggaaACTAGTggaggaacaaaggaggaagg	30	72.6	20		

Table 25 Primers used to amplify ECR1-*LSAMP*, ECR2-*LSAMP* and ECR3-*LSAMP* from whole-chick genomic DNA.

The Primer3-generated forward primers are highlighted in green, the reverse primer highlighted in red. The NotI (green capital letters) and SpeI (red capital letters) sequences were inserted to allow directional cloning. The short sequences distal to the main sequence were appended to allow the restriction enzymes to function.

The full length of the primer, and the total length, Tm of the total primer as calculated by the oligonucleotide manufacturer (Sigma), the primer length, product size and melting temperature of the highlighted sequences generated by Primer3 are also shown.

7.10 Primers ordered for to PCR-amplify second tranche ECRs

A (ECR1)

GC|GGCCGCatctgaacgttgccctgtcacggaggcacagcacaaggttgcagcaggccctggataagaaagcactgaggaatggg
aagttggctactcacCTGTAGTGCAAAACAACCTCATTGATACGCAGGAGACTTCAGTCTCGGTGACTTTCAATGTCTTCCACAGC
ATCAAACTGATATTTCAAAGAGATTAAGAACTGAAGGGAACATATGATATCCATTTAAAGCCAAGACTTCAGATTTAATGAAGGCT
GTATCAGAATGAGATATGATTCAGATATCAAATTACATTTGTTTAGCCCAACACAGCTTAATAAATCACAGCAACATTTAATAAAT
ATTTATCTTCTTTTAATCACCAGTTAGTGCTTTTTCTGATCTCTCCATTCCCTTAATGGCTCACTCCTGTATTTATTAAATGTGAT
TTCCATTTCTctaataatcaatacatctttttatttttataatttgggcaGctttactgaactA|CTAGT

B (ECR1)

GGTGGC|GGCCGCatctgaacgttgccctgtcacggaggcacagcacaaggttgcagcaggccctggataagaaagcactgaggaa
tgggaagttggctactcacCTGTAGTGCAAAACAACCTCATTGATACGCAGGAGACTTCAGTCTCGGTGACTTTCAATGTCTTCCA
CAGCATCAAACTGATATTTCAAAGAGATTAAGAACTGAAGGGAACATATGATATCCATTTAAAGCCAAGACTTCAGATTTAATGAA
GGCTGTATCAGAATGAGATATGATTCAGATATCAAATTACATTTGTTTAGCCCAACACAGCTTAATAAATCACAGCAACATTTAAT
AAATATTTATCTTCTTTTAATCACCAGTTAGTGCTTTTTCTGATCTCTCCATTCCCTTAATGGCTCACTCCTGTATTTATTAAATG
TGATTTCCATTTCTctaataatcaatacatctttttatttttataatttgggcagctttactgaactA|*CTAGTGGATCCCCCGG*
GCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTAGCCTGCAGGTCGAGGAGCGCAGCCTTCCAGAAGCA
GAGCGCGGCGCCATGGGGA

C (ECR2)

GC|GGCCGCagtccagctgccaacagaatggTACCAATGATAAAATGTTATTGTTGTCACTGTCAGGCATTTGGTCTTTATCGGTC
TACTCATGACAGCCCTTGTTAGTCAGGCAGGTCCTGGGACAACTTCAAAGGGAGACCTTCTCATCCAATGACAaacatagcaacct
acttgacttctctgatcctcttatttccttcaggccttctgcA|CTAGT

D (ECR2)

AGTGACTATGCGGGTGGC|GGCCGCAGTCCAGCTGCCAACAGAATggTACCAATGATAAAATGTTATTGTTGTCACTGTCAGGCAT
TTGGTCTTTATCGGTCTACTCATGACAGCCCTTGTTAGTCAGGCAGGTCCTGGGACAACTTCAAAGGGAGACCTTCTCATCCAATG
ACAaacatagcaacctacttgacttctctgatcctcttATTTCCTTCAGGCCTTCTGCA|*CTAGTGGATCCCCCGGGCTGGGCATA*
AAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTAGCCTGCAGGTCGAGGAGCGCAGCCTTCCAGAAGCAGAGCGCGGCG

E (ECR3)

GC|GGCCGCctgtctgcctgtctgtctgcaggggtatggtgtaagggtgggtctttcccgctatagtttttggtTGGAAGGAGAC
TCTGAGAGGGGTGTAGAAAAGGAAGGAGATAACCCCCAGGAAAGTCAAACTATCTATTTATTTAGTTATGTGTTTTGGGTTGCGTT
TTGCATTAATCGCTGCCACACACAGGAGCCCTGTAATTTATTATTGGTGATAAGTGTGGACAGTAGCTATTATTTTTGCTCCCTGT
TCTAAAACGTGGCAAGCAGGGACACTTAGATTTATGGAACATGCATTATTCAAATCTCAAGGCCCGATGGCTCAGCATGAACAACC
CCAGCTATTTCCAGAACCTCTcagtccttcctcctttgttcctccA|CTAGT

F (ECR3)

GACGGGGACTCCTCGTCGGGTGGC|GGCCGCCTGTCTGCCTGTCTGTCTGCaggggtatggtgtaagggtgggtctttcccgcta
tagtttttggtTGAAAGGAGACTCTGAGAGGGGTGTAGAAAAGGAAGGAGATAACCCCCAGGAAAGTCAAACTATCTATTTATTTA
GTTATGTGTTTTGGGTTGCGTTTTGCATTAATCGCTGCCACACACAGGAGCCCTGTAATTTATTATTGGTGATAAGTGTGGACAGT
AGCTATTATTTTTGCTCCCTGTTCTAAAACGTGGCAAGCAGGGACACTTAGATTTATGGAACATGCATTATTCAAATCTCAAGGCC
CGATGGCTCAGCATGAACAACCCCAGCTATTTCCAGAACCTCTcagtCCTTCCTCCTTTGTTCCTCCA|*CTAGTGGATCCCCCGGG*
CTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTAGCCTGCAGGTCGAGGAGCGCAGCCTTCCAGAAGCAG
AGCGCGGCGCCATGGGGGATCCCGTCGTTT

Figure 78. Sequencing results showing ECRs have been cloned into p1229.

(A) ECR1 insert bracketed by NotI and SpeI sequences.

(B) Results of the sequencing of clone ABM6 from the T7 promoter showing the ECR1 sequence cloned into p1229 in the forward direction.

(C) ECR2 insert bracketed by NotI and SpeI sequences.

(D) Results of the sequencing of clone 5 from the T7 promoter showing the ECR2 sequence cloned into p1229 in the forward direction.

E) ECR3 insert bracketed by NotI and SpeI sequences.

(F) Results of the sequencing of clone F from the T7 promoter showing the ECR3 sequence cloned into p1229 in the forward direction. There is one G:A substitution (pink highlight) which could be a sequencing error.

KEY:

Yellow highlight; PCR-amplified sequence within primer regions. Upper case; ECR sequence. Lower case; regions flanking ECR.

Green highlight; Forward primer's annealing sequence. Red highlight; Reverse primer's annealing sequence.

NotI cutting sequence; GC|GGCCGC, cuts at |. SpeI cutting sequence; A|CTAGT, cuts at |.

Un-highlighted italics; p1229 reporter vector sequence.

7.11 Alignment between predicted and sequenced *LSAMP* ECR1,2,3 inserts

A

```
>lcl|19999
Length=619
 Score =  922 bits (499),  Expect = 0.0
 Identities = 499/499 (100%), Gaps = 0/499 (0%)
 Strand=Plus/Plus

Query  1    GCGGCCGCATCTGAACGTTGCCCTGTCACGGAGGCACAGCACAAGGTTGCAGCAGGCCCT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5    GCGGCCGCATCTGAACGTTGCCCTGTCACGGAGGCACAGCACAAGGTTGCAGCAGGCCCT  64

Query  61   GGATAAGAAAGCACTGAGGAATGGGAAGTTGGCTACTCACCTGTAGTGCAAAACAACCTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  65   GGATAAGAAAGCACTGAGGAATGGGAAGTTGGCTACTCACCTGTAGTGCAAAACAACCTC  124

Query  121  ATTGATACGCAGGAGACTTCAGTCTCGGTGACTTTCAATGTCTTCCACAGCATCAAACTG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  125  ATTGATACGCAGGAGACTTCAGTCTCGGTGACTTTCAATGTCTTCCACAGCATCAAACTG  184

Query  181  ATATTTCAAAGAGATTAAGAACTGAAGGGAACATATGATATCCATTTAAAGCCAAGACTT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  185  ATATTTCAAAGAGATTAAGAACTGAAGGGAACATATGATATCCATTTAAAGCCAAGACTT  244

Query  241  CAGATTTAATGAAGGCTGTATCAGAATGAGATATGATTCAGATATCAAATTACATTTGTT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  245  CAGATTTAATGAAGGCTGTATCAGAATGAGATATGATTCAGATATCAAATTACATTTGTT  304

Query  301  TAGCCCAACACAGCTTAATAAATCACAGCAACATTTAATAAATATTTATCTTCTTTTAAT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  305  TAGCCCAACACAGCTTAATAAATCACAGCAACATTTAATAAATATTTATCTTCTTTTAAT  364

Query  361  CACCAGTTAGTGCTTTTTCTGATCTCTCCATTCCCTTAATGGCTCACTCCTGTATTTATT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  365  CACCAGTTAGTGCTTTTTCTGATCTCTCCATTCCCTTAATGGCTCACTCCTGTATTTATT  424

Query  421  AAATGTGATTTCCATTTCTCTAATAATCAATACATCttttttatttttttATAATTTGGGCA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  425  AAATGTGATTTCCATTTCTCTAATAATCAATACATCTTTTTATTTTTTATAATTTGGGCA  484

Query  481  GCTTTACTGAACTACTAGT  499
            ||||||||||||||||||
Sbjct  485  GCTTTACTGAACTACTAGT  503
```

B

```
>lcl|54795

Length=342

 Score =  405 bits (219),  Expect = 6e-118
 Identities = 219/219 (100%), Gaps = 0/219 (0%)
 Strand=Plus/Plus

Query  1    GCGGCCGCAGTCCAGCTGCCAACAGAATGGTACCAATGATAAAATGTTATTGTTGTCACT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  17   GCGGCCGCAGTCCAGCTGCCAACAGAATGGTACCAATGATAAAATGTTATTGTTGTCACT  76

Query  61   GTCAGGCATTTGGTCTTTATCGGTCTACTCATGACAGCCCTTGTTAGTCAGGCAGGTCCT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  77   GTCAGGCATTTGGTCTTTATCGGTCTACTCATGACAGCCCTTGTTAGTCAGGCAGGTCCT  136

Query  121  GGGACAACTTCAAAGGGAGACCTTCTCATCCAATGACAAACATAGCAACCTACTTGACTT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  137  GGGACAACTTCAAAGGGAGACCTTCTCATCCAATGACAAACATAGCAACCTACTTGACTT  196

Query  181  CTCTGATCCTCTTATTTCCTTCAGGCCTTCTGCACTAGT  219
            |||||||||||||||||||||||||||||||||||||||
Sbjct  197  CTCTGATCCTCTTATTTCCTTCAGGCCTTCTGCACTAGT  235
```

C

```
>lcl|27485
Length=543

 Score =  719 bits (389),  Expect = 0.0
 Identities = 391/392 (99%), Gaps = 0/392 (0%)
 Strand=Plus/Plus

Query  3    GGCCGCCTGTCTGCCTGTCTGTCTGCAGGGGGTATGGTGTAAGGGTGGGTCTTTCCCGCT  62
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  24   GGCCGCCTGTCTGCCTGTCTGTCTGCAGGGGGTATGGTGTAAGGGTGGGTCTTTCCCGCT  83

Query  63   ATAGTTTTTGGTTGGAAGGAGACTCTGAGAGGGGTGTAGAAAAGGAAGGAGATAACCCCC  122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  84   ATAGTTTTTGGTTGAAAGGAGACTCTGAGAGGGGTGTAGAAAAGGAAGGAGATAACCCCC  143

Query  123  AGGAAAGTCAAACTATCTATTTATTTAGTTATGTGTTTTGGGTTGCGTTTTGCATTAATC  182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  144  AGGAAAGTCAAACTATCTATTTATTTAGTTATGTGTTTTGGGTTGCGTTTTGCATTAATC  203

Query  183  GCTGCCACACACAGGAGCCCTGTAATTTATTATTGGTGATAAGTGTGGACAGTAGCTATT  242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  204  GCTGCCACACACAGGAGCCCTGTAATTTATTATTGGTGATAAGTGTGGACAGTAGCTATT  263

Query  243  ATTTTTGCTCCCTGTTCTAAAACGTGGCAAGCAGGGACACTTAGATTTATGGAACATGCA  302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  264  ATTTTTGCTCCCTGTTCTAAAACGTGGCAAGCAGGGACACTTAGATTTATGGAACATGCA  323

Query  303  TTATTCAAATCTCAAGGCCCGATGGCTCAGCATGAACAACCCCAGCTATTTCCAGAACCT  362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  324  TTATTCAAATCTCAAGGCCCGATGGCTCAGCATGAACAACCCCAGCTATTTCCAGAACCT  383

Query  363  CTCAGTCCTTCCTCCTTTGTTCCTCCACTAGT  394
            ||||||||||||||||||||||||||||||||
Sbjct  384  CTCAGTCCTTCCTCCTTTGTTCCTCCACTAGT  415
```

Figure 79 Alignments between predicted and sequenced (from T7 promoter) of (A) ECR1 (clone ABM6) (B) ECR2 (clone 5) (C) ECR3 (clone F) sequences cloned into the p1229 construct.

7.12 PCR sequences ordered to amplify second tranche of ECRs

ECR Name	Direction	Sequence ordered
Grp1Ecr2	F	TGAGACTGACCAGACACAAAGG
	R	CCATGGTGAGACAGGGTTTT
Grp1Ecr3	F	CCTCATTAACACACCGCTCA
	R	TGCTGAGTTGCAAAGGACAG
Grp2Ecr4	F	TTTGTTGTGAGAGGCTGTGC
	R	TCCCTTAATCTGTTGTGTTCCA
Grp2Ecr5	F	TTCTCCCATACATCCTTGTTGTT
	R	CCAGACAGGAGGAAAGAAGC
Grp2Ecr6	F	TGACAAACTCAAACCTGAAGC
	R	GCGTTTGTCCAAGGAAGAAG
Grp3Ecr3	F	CAGGGAATGTGAGGAAGAGG
	R	ACCAGCCTCTACCCCTTTGT
Grp3Ecr4	F	CAGAGACTTTGGGGATCACTG
	R	AAACCAGACAAGGGGCAAC
Grp3Ecr5	F	CGTGTGGAAAGAGAGGGATG
	R	GCTGTCAGCTAGGGACAAATG

Table 26 Primers ordered for second tranche of ECRs. As directional cloning was not being performed for these ECRs the NotI and SpeI sites were not required.

7.14 List of TFBS located within the ECRs

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$ACAAT_B	ACAAT	HISTONE	Y								Y	Y
V$AFP1_Q6	AFP1	HOX		Y			Y					
V$AHR_Q5	AhR	BHLH								YY		
V$AHRARNT_02	AhR:Arnt	BHLH		Y								
V$AIRE_02	AIRE	ZFPHD							Y			
V$ALPHACP1_01	alpha-CP1	HISTONE									Y	YY
V$ALX4_01	Alx-4	HOX						Y				
V$AP1_01	AP-1	BZIP				Y						
V$AP1_Q2	AP-1	BZIP						Y				
V$AP1_Q4	AP-1	BZIP					Y	Y				
V$AP1_Q6	AP-1	BZIP						Y				
V$AP1_Q6_01	AP-1	BZIP					Y					
V$AP1FJ_Q2	AP-1	BZIP						Y				
V$APOLYA_B	APOLYA	PolyA?							Y			
V$AREB6_01	AREB6	ZFC2H2										Y
V$ATF4_Q2	ATF4	BZIP						Y				
V$BACH1_01	Bach1	BZIP								Y		
V$BARBIE_01	Barbie Box	PLZF	Y									
V$BLIMP1_Q6	Blimp-1	ZFC2H2						Y				

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$CEBP_01	C/EBP	BZIP						Y	Y			
V$CEBP_Q2_01	C/EBP	BZIP									Y	
V$CEBP_Q3	C/EBP	BZIP									Y	
V$CHX10_01	C/EBP	BZIP					Y	Y				
V$CEBP_Q2	C/EBP alpha	BZIP		Y								
V$CEBPA_01	C/EBPalpha	BZIP									Y	
V$CEBPB_02	C/EBPbeta	BZIP									Y	
V$CEBPGAMMA_Q6	C/EBPgamma	BZIP	Y					Y				
V$CART1_01	Cart-1	HOX						YY	Y			
V$CAAT_01	CCAAT box	HISTONE									Y	Y
V$CDC5_01	Cdc5	MYB			Y				Y			
V$CDP_01	CDP	HOX				Y			Y			
V$CDP_02	CDP	HOX							Y			
V$CDPCR3_01	CDP CR3	HOX				Y						
V$CDPCR3HD_01	CDP CR3+HD	HOX	Y							Y		
V$CDX_Q5	Cdx	HOX				Y						
V$CDX2_Q5	Cdx-2	HOX					Y					
V$CDXA_01	CdxA	HOX				Y						
V$CDXA_02	CdxA	HOX								Y		
V$CLOX_01	Clox	HOX							Y			
V$CMAF_01	c-Maf	BZIP						Y				
V$MYB_Q6	c-Myb	MYB						Y				
V$COMP1_01	COMP1	--									Y	

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$CP2_01	CP2	GRAINY					Y					
V$CREB_02	CREB	BZIP						Y				
V$CREB_Q3	CREB	BZIP							Y			
V$CRX_Q4	Crx	HOX	Y		Y							
V$DEAF1_01	DEAF1	SAND				Y						
V$DTYPEPA_B	D-Type LTRs	D-Type LTRs									Y	
V$E2_Q6_01	E2	E2							Y			
V$E2F_02	E2F	E2						Y				
V$E2F_03	E2F	E2						Y			Y	
V$E2F_Q3	E2F	E2F						Y				
V$E2F_Q4	E2F	E2F						Y				
V$E2F_Q4_01	E2F	E2F						Y				
V$E2F_Q6	E2F	E2F						Y				
V$E2F_Q6_01	E2F	E2F						Y				
V$E2F1_Q3_01	E2F	E2F									Y	
V$E2F1_Q6_01	E2F	E2F						Y				
V$E2F1DP1_01	E2F-1:DP-1	E2F						Y				
V$E2F1DP2_01	E2F-1:DP-2	E2F						Y				

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$E2F4DP1_01	E2F-4:DP-1	E2F						Y				
V$E2F4DP2_01	E2F-4:DP-2	E2F						Y				
V$E47_01	E47	BHLH						Y				
V$E4BP4_01	E4BP4	BZIP									Y	YY
V$EN1_01	EN1	HOX		Y			Y	Y	Y			
V$ER_Q6_02	ER	ZFC4-NR					YY					
V$ERR1_Q2	ERR1	ZFC4-NR					Y					Y
V$EVI1_01	Evi-1	ZFC2H2	YY									
V$EVI1_03	Evi-1	ZFC2H2	Y				Y					
V$EVI1_06	Evi-1	ZFC2H2			Y							
V$FAC1_01	FAC1	ZFPHD				Y						
V$FOXJ2_01	FOXJ2	FORKHEAD			Y	Y						
V$FOXJ2_02	FOXJ2	FORKHEAD						Y				
V$FOXM1_01	FOXJ2	FORKHEAD				Y						
V$FOXM1_01	FOXm1	FORKHEAD							Y			
V$FREAC3_01	Freac-3	FORKHEAD				Y						
V$FREAC7_01	Freac-7	FORKHEAD					Y		Y			
V$FXR_Q3	FXR/RXR-alpha	ZFC4-NR	Y									

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$GATA_Q6	GATA	ZFGATA				Y						
V$GATA1_02	GATA-1	ZFGATA				Y		Y				
V$GATA1_03	GATA-1	ZFGATA						Y				
V$GATA1_06	GATA-1	ZFGATA						Y				
V$GATA2_02	GATA-2	ZFGATA						Y				
V$GATA2_03	GATA-2	ZFGATA			Y							
V$GATA3_01	GATA-3	ZFGATA						Y				
V$GATA3_02	GATA-3	ZFGATA						Y				
V$GATA4_Q3	GATA-4	ZFGATA						Y				
V$GATA_C	GATA-X	ZFGATA			Y							
V$GCM_Q2	GCM	GCM						Y				
V$GCNF_01	GCNF	ZFC4-NR						Y				
V$GEN_INI_B	GEN_INI	GENINI			Y							
V$GEN_INI3_B	GEN_INI	GENINI			Y							
V$GFI1_Q6	Gfi1	ZFC2H2			Y							
V$GFI1_01	Gfi-1	ZFC2H2			Y				Y			
V$GFI1B_01	Gfi1b	ZFC2H2			Y							

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$HELIOSA_01	Helios A	ZFC2H2							Y			
V$HEN1_01	HEN1	BHLH						Y				
V$HEN1_02	HEN1	BHLH						YY				
V$HFH3_01	HFH3 (FOXI1)	FORKHEAD		Y		Y			Y			
V$HLF_01	HLF	BZIP										Y
V$HNF3_Q6	HNF3	FORKHEAD				Y						
V$HNF3_Q6_01	HNF3	FORKHEAD				YY						
V$HNF3ALPHA_Q6	HNF3	FORKHEAD		Y		Y						
V$HNF3B_01	HNF3	FORKHEAD		Y		Y						
V$HNF4_01_B	HNF4	FORKHEAD						Y				
V$HNF4_Q6	HNF4	FORKHEAD				Y			Y			Y
V$HOXA7_01	HOXA7	HOX								Y		
V$ICSBP_Q6	ICSBP	IRF				Y						
V$IRF1_Q6	IRF-1	IRF		Y								
V$ISRE_01	ISGF-3	IRF				Y						
V$LEF1_Q2	LEF1	HMG			Y		Y		Y			
V$LEF1_Q2_01	LEF1	HMG			Y		Y		Y			
V$LEF1B_01	LEF1	HMG			Y		Y		Y			
V$LPOLYA_B	Lentiviral Poly A	PolyA?								Y		
V$LHX3_01	Lhx3	HOX	Y		Y		Y					
V$LMO2COM_02	Lmo2 complex	ZFGATA				Y		Y				
V$LUN1_01	LUN-1	ZFRING				Y						
V$MAF_Q6_01	MAF	BZIP				Y		Y				

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$MEF2_02	MEF-2	MADS								Y		
V$MEF2_03	MEF-2	MADS								Y		
V$MEF2_Q6_01	MEF-2	MADS								Y		
V$MMEF2_Q6	MEF-2	MADS					Y					
V$MIF1_01	MIF-1	RFX					Y					
V$MRF2_01	MRF-2	ARID							Y			
V$MSX1_01	Msx-1	HOX		Y			Y	Y	Y			
V$MUSCLE_INI_B	Muscle initiator	GENINI				Y						
V$MINI19_B	Muscle initiator sequences-19	GENINI				Y						
V$MINI20_B	Muscle initiator sequences-20	GENINI				Y						Y
V$NCX_01	Ncx	HOX					Y					
V$NFE2_01	NF-E2	BZIP				Y		Y				
V$NFY_01	NF-Y	HISTONE									Y	Y
V$NFY_C	NF-Y	HISTONE									Y	Y
V$NFY_Q6	NF-Y	HISTONE				Y					Y	YY
V$NFY_Q6_01	NF-Y	HISTONE									Y	
V$NKX22_01	Nkx2-2	HOX									Y	
V$NKX25_01	Nkx2-5	HOX						Y				
V$NKX25_02	Nkx2-5	HOX			Y			Y			Y	
V$NKX25_Q5	Nkx2-5	HOX						Y				
V$NKX25B_1	Nkx2-5	HOX						Y				

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$NKX25L_01	Nkx2-5	HOX			Y			Y		Y	Y	
V$NKX62_Q2	Nkx6-2	HOX		Y	YY							
V$NRF2_Q4	Nrf-2	BZIP				Y						
V$OCT_Q6	Octamer	HOX			Y							
V$OCT_C	OCT-x	HOX									Y	
V$OTX_Q1	OTX	HOX	Y								Y	
V$P53_DECAMER_Q2	p53 decamer	P53							Y			
V$PAX_Q6	Pax	HOX						Y				
V$PAX2_01	Pax-2	HOX								Y		
V$PAX2_02	Pax-2	HOX							Y			
V$PAX6_01	Pax-6	HOX						YY				
V$PAX8_B	Pax-8	HOX						Y				
V$PBX_Q3	Pbx	HOX	Y							Y		
V$PBX1_01	Pbx-1	HOX							Y			
V$PBX1_02	Pbx-1	HOX							Y	Y		
V$PBX1_03	Pbx-1	HOX	Y						Y	YY		
V$IPF1_Q4_01	PDX1 (ipf1)	HOX			Y				Y			
V$POU1F1_Q6	POU1F1	HOX					Y					
V$PPARG_03	PPAR	ZFC4-NR					Y	Y				
V$PPAR_DR1_Q2	PPAR direct repeat 1	ZFC4-NR						Y				
V$DR1_Q3	PPAR, HNF-4, COUP, RAR	ZFC4-NR						Y				
V$PPARG_01	PPARgamma:RXRalpha	ZFC4-NR						Y				

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$PXR_Q2	PXR, CAR, LXR, FXR	ZFC4-NR							Y			
V$R_01	R	BHLH							Y			
V$E2F1DP1RB_01	Rb:E2F-1:DP-1	E2F						Y				
V$POLY_C	Retroviral Poly A	PolyA?					Y		Y	YY		
V$RORA1_01	RORalpha1	ZFC4-NR					Y					Y
V$RSRFC4_01	RSRFC4	MADS								Y		
V$RSRFC4_Q2	RSRFC4	MADS								Y		
V$S8_01	S8	HOX			Y				Y			
V$SF1_Q6	SF1	ZFC4-NR					Y	Y				
V$SMAD_Q6	SMAD	SMAD							Y			
V$SMAD_Q6_01	SMAD	SMAD	Y									
V$SMAD3_Q6	SMAD	SMAD				Y						
V$SMAD4_Q6	SMAD	SMAD						Y				
V$SRF_C	SRF	MADS									Y	
V$SRF_Q6	SRF	MADS						Y				
V$STAT1_02	STAT1	STAT		Y								
V$STAT1_03	STAT1	STAT			Y							
V$STAT4_01	STAT4	STAT			Y							
V$STAT5A_04	STAT5	STAT									Y	
V$STAT6_01	STAT6	STAT				Y						

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$STRA13_01	Stra13	BHLH										Y
V$TATA_01	TATA	TBP	Y							Y		
V$TATA_C	TATA	TBP	Y									
V$TBP_01	TBP	TBP			Y					Y		
V$TBP_Q6	TBP	TBP	Y									
V$TBX5_02	TBX5	TBX							Y			
V$TCF11_01	TCF11	BZIP					Y					
V$TCF11MAFG_01	TCF11	BZIP					Y					
V$TCF4_Q5	TCF-4	HMG					Y					
V$TFE_Q6	TFE	BHLH										Y
V$TST1_01	Tst-1	HOX						Y				
V$TITF1_Q3	TTF1 (Nkx2-1)	HOX						Y				
V$TTF1_Q6	TTF-1 (Nkx2-1)	HOX				Y					Y	
V$USF_Q6	USF	BHLH					Y					
V$VBP_01	VBP	BZIP										Y
V$DR3_Q4	VDR, CAR, PXR	ZFC4-NR		Y								
V$T3R_01	v-ErbA	ZFC4-NR					Y					Y
V$VMYB_01	v-Myb	MYB			Y							
V$VMYB_02	v-Myb	MYB						Y				
V$XBP1_01	XBP-1	BZIP							Y			
V$XFD2_01	XBP-1	BZIP			Y							
V$XVENT1_01	Xvent-1	HOX									Y	

TFBS Matrix Name	Transcription Factor Name	TF class	G1E2	G1E3	G2E1 ECR1	G2E3 ECR3	G2E4	G2E5	G2E6	G3E3	G3E4	G3E5
V$YY1_01	YY1	ZFC2H2						Y	Y			
V$OCT1		HOX		Y								
V$OCT1_Q5_01	Oct-1	HOX			Y							
V$OCT1_Q6	Oct-1	HOX			Y							
V$OCT1_01	Oct-1	HOX					Y					
V$OCT1_03	Oct-1	HOX						Y			Y	Y
V$OCT1_05	Oct-1	HOX			Y							
V$OCT1_B	Oct-1	HOX				Y						
V$OCT4_01	Oct-4 (POU5F1)	HOX						Y		Y		
V$OCT4_02	Oct-4 (POU5F1)	HOX						Y		Y		

Table 27 List of TFBS in the ECRs that were tested for enhancer activity. Red columns indicate ECRs that were not shown to drive expression in the reporter plasmid. TFBS in red text and salmon coloured rows indicate TFBS that are solely in those negative ECRs.

TF Class	E	B	N
ARID	0	0	1
GRAINY	0	0	1
P53	0	0	1
RFX	0	0	1
TBX	0	0	1
PolyA	0	0	3
HMG	0	0	4
GCM	1	0	0
PLZF	1	0	0
SAND	1	0	0
Unclassified	1	0	0
ZFRING	1	0	0
D-Type LTRs	1	0	1
ZFPHD	1	0	1
E2	2	0	0
TBP	2	1	1
MYB	2	0	2
MADS	2	0	6
IRF	3	0	0
SMAD	3	0	1
GENINI	3	0	2
STAT	3	0	2
ZFC2H2	3	2	5
ZFC4-NR	5	5	2
BHLH	6	0	3
HISTONE	7	0	0
FORKHEAD	9	1	2
ZFGATA	9	0	2
E2F	12	0	0
BZIP	19	3	7
HOX	20	15	18
Totals	117	27	67

Table 28 TFBSs within all *LSAMP*-related ECRs tested in this thesis.
Key: E = TFBSs located in ECRs shown to have only enhancer activity
B = TFBSs located in ECRs that had both enhancer and no enhancer activity in this tissue and at this stage of development
N = TFBSs located in ECRs that only showed no enhancer activity in this tissue and at this stage of development

BIBLIOGRAPHY

AKEEL, M., MCNAMEE, C. J., YOUSSEF, S. & MOSS, D. 2011. DIgLONs inhibit initiation of neurite outgrowth from forebrain neurons via an IgLON-containing receptor complex. *Brain Research,* 1374, 27-35.
ALBERTS, B. 2002. *Molecular biology of the cell,* New York, Garland Science.
ALBERTS, B., WILSON, J. H. & HUNT, T. 2008. *Molecular biology of the cell,* New York, Garland Science.
ALCARAZ-PEREZ, F., MULERO, V. & CAYUELA, M. 2008. Application of the dual-luciferase reporter assay to the analysis of promoter activity in Zebrafish embryos. *BMC Biotechnology,* 8, 81.
ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215, 403-10.
ALWAHHABI, F. 2003. Anxiety symptoms and generalized anxiety disorder in the elderly: a review. *Harvard review of psychiatry,* 11, 180-93.
AMERICAN PSYCHIATRIC ASSOCIATION 2000. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR.*
ANDREW, S. E., GOLDBERG, Y. P., KREMER, B., TELENIUS, H., THEILMANN, J., ADAM, S., STARR, E., SQUITIERI, F., LIN, B., KALCHMAN, M. A. & ET AL. 1993. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet,* 4, 398-403.
ANGERS-LOUSTAU, A., COTE, J. F. & TREMBLAY, M. L. 1999. Roles of protein tyrosine phosphatases in cell migration and adhesion. *Biochem Cell Biol,* 77, 493-505.
ARBER, S., HAN, B., MENDELSOHN, M., SMITH, M., JESSELL, T. M. & SOCKANATHAN, S. 1999. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron,* 23, 659-74.
ARUNACHALAM, M., JAYASURYA, K., TOMANCAK, P. & OHLER, U. 2010. An alignment-free method to identify candidate orthologous enhancers in multiple Drosophila genomes. *Bioinformatics,* 26, 2109-2115.
ASANO, T., KATAGIRI, H., TAKATA, K., LIN, J. L., ISHIHARA, H., INUKAI, K., TSUKUDA, K., KIKUCHI, M., HIRANO, H., YAZAKI, Y. & ET AL. 1991. The role of N-glycosylation of GLUT1 for glucose transport activity. *The Journal of biological chemistry,* 266, 24632-6.
BADER, G. D. & HOGUE, C. W. 2000. BIND--a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics,* 16, 465-77.
BAHR, J. M. 2008. The Chicken as a Model Organism. *In:* CONN, P. M. (ed.) *Sourcebook of Models for Biomedical Research.* Totowa, NJ.: Humana Press Inc.
BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27, 299-308.
BARCLAY, A. N. 1999. Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America,* 96, 14672-4.
BASSON, M. D. 2008. An intracellular signal pathway that regulates cancer cell adhesion in response to extracellular forces. *Cancer Res,* 68, 2-4.

BATEMAN, A., EDDY, S. R. & CHOTHIA, C. 1996. Members of the immunoglobulin superfamily in bacteria. *Protein science : a publication of the Protein Society,* 5, 1939-41.
BATEMAN, J. R., JOHNSON, J. E. & LOCKE, M. N. 2012. Comparing enhancer action in cis and in trans. *Genetics,* 191, 1143-55.
BAUER, F., ELBERS, C. C., ADAN, R. A., LOOS, R. J., ONLAND-MORET, N. C., GROBBEE, D. E., VAN VLIET-OSTAPTCHOUK, J. V., WIJMENGA, C. & VAN DER SCHOUW, Y. T. 2009. Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference. *Am J Clin Nutr,* 90, 951-9.
BAUMEISTER, H. & HARTER, M. 2007. Prevalence of mental disorders based on general population surveys. *Social psychiatry and psychiatric epidemiology,* 42, 537-46.
BECKER, J. W., ERICKSON, H. P., HOFFMAN, S., CUNNINGHAM, B. A. & EDELMAN, G. M. 1989. Topology of cell adhesion molecules. *Proceedings of the National Academy of Sciences of the United States of America,* 86, 1088-92.
BECKER, K. G., JEDLICKA, P., TEMPLETON, N. S., LIOTTA, L. & OZATO, K. 1994. Characterization of hUCRBP (YY1, NF-E1, delta): a transcription factor that binds the regulatory regions of many viral and cellular genes. *Gene,* 150, 259-66.
BECKERLE, M. C. 2001. *Cell adhesion,* Oxford, Oxford University Press.
BEEKMAN, A. T., BREMMER, M. A., DEEG, D. J., VAN BALKOM, A. J., SMIT, J. H., DE BEURS, E., VAN DYCK, R. & VAN TILBURG, W. 1998. Anxiety disorders in later life: a report from the Longitudinal Aging Study Amsterdam. *International journal of geriatric psychiatry,* 13, 717-26.
BEGGS, H. E., BARAGONA, S. C., HEMPERLY, J. J. & MANESS, P. F. 1997. NCAM140 interacts with the focal adhesion kinase p125(fak) and the SRC-related tyrosine kinase p59(fyn). *J Biol Chem,* 272, 8310-9.
BEHAN, A. T., BYRNE, C., DUNN, M. J., CAGNEY, G. & COTTER, D. R. 2009. Proteomic analysis of membrane microdomain-associated proteins in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder reveals alterations in LAMP, STXBP1 and BASP1 protein expression. *Mol Psychiatry,* 14, 601-13.
BELL, G. W., YATSKIEVYCH, T. A. & ANTIN, P. B. 2004. GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos. *Developmental dynamics : an official publication of the American Association of Anatomists,* 229, 677-87.
BIEDERER, T. 2006. Bioinformatic characterization of the SynCAM family of immunoglobulin-like domain-containing adhesion molecules. *Genomics,* 87, 139-50.
BIEDERER, T., SARA, Y., MOZHAYEVA, M., ATASOY, D., LIU, X., KAVALALI, E. T. & SUDHOF, T. C. 2002. SynCAM, a synaptic adhesion molecule that drives synapse assembly. *Science,* 297, 1525-31.
BIER, E. & MCGINNIS, W. 2004. Model Organisms in the study of Development and Disease. *In:* EPSTEIN, C. J., ERICKSON, R. P. & WYNSHAW-BORIS, A. (eds.) *Inborn Errors of Development: The*

Molecular Basis of Clinical Disorders of Morphogenesis. Oxford University Press, USA.

BOBADILLA, J. L., MACEK, M., JR., FINE, J. P. & FARRELL, P. M. 2002. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. *Hum Mutat,* 19, 575-606.

BORK, P., HOLM, L. & SANDER, C. 1994. The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of molecular biology,* 242, 309-20.

BOSHART, M., WEBER, F., JAHN, G., DORSCH-HASLER, K., FLECKENSTEIN, B. & SCHAFFNER, W. 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell,* 41, 521-30.

BOUYAIN, S. & WATKINS, D. J. 2010. Identification of tyrosine phosphatase ligands for contactin cell adhesion molecules. *Commun Integr Biol,* 3, 284-6.

BREEN, G., WEBB, B. T., BUTLER, A. W., VAN DEN OORD, E. J., TOZZI, F., CRADDOCK, N., GILL, M., KORSZUN, A., MAIER, W., MIDDLETON, L., MORS, O., OWEN, M. J., COHEN-WOODS, S., PERRY, J., GALWEY, N. W., UPMANYU, R., CRAIG, I., LEWIS, C. M., NG, M., BREWSTER, S., PREISIG, M., RIETSCHEL, M., JONES, L., KNIGHT, J., RICE, J., MUGLIA, P., FARMER, A. E. & MCGUFFIN, P. 2011. A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. *The American journal of psychiatry,* 168, 840-7.

BRESLAUER, K. J., FRANK, R., BLÖCKER, H. & MARKY, L. A. 1986. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences,* 83, 3746-3750.

BRITTON, J. C., LISSEK, S., GRILLON, C., NORCROSS, M. A. & PINE, D. S. 2011. Development of anxiety: the role of threat appraisal and fear learning. *Depression and anxiety,* 28, 5-17.

BRUMMENDORF, T. & RATHJEN, F. G. 1995. Cell adhesion molecules 1: immunoglobulin superfamily. *Protein Profile,* 2, 963-1108.

BRUMMENDORF, T., SPALTMANN, F. & TREUBERT, U. 1997. Cloning and characterization of a neural cell recognition molecule on axons of the retinotectal system and spinal cord. *Eur J Neurosci,* 9, 1105-16.

BUCK, C. A. 1992. Immunoglobulin superfamily: structure, function and relationship to other receptor molecules. *Semin Cell Biol,* 3, 179-88.

BURRIDGE, K., SASTRY, S. K. & SALLEE, J. L. 2006. Regulation of cell adhesion by protein-tyrosine phosphatases. I. Cell-matrix adhesion. *J Biol Chem,* 281, 15593-6.

BURT, D. W., BRULEY, C., DUNN, I. C., JONES, C. T., RAMAGE, A., LAW, A. S., MORRICE, D. R., PATON, I. R., SMITH, J., WINDSOR, D., SAZANOV, A., FRIES, R. & WADDINGTON, D. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature,* 402, 411-3.

BURT, D. W., CARRE, W., FELL, M., LAW, A. S., ANTIN, P. B., MAGLOTT, D. R., WEBER, J. A., SCHMIDT, C. J., BURGESS, S. C. & MCCARTHY, F. M. 2009. The Chicken Gene Nomenclature Committee report. *BMC Genomics,* 10 Suppl 2, S5.

BUTLER, H. & JUURLINK, B. H. J. 1987. *An atlas for staging mammalian and chick embryos,* Boca Raton, Fla., CRC Press.

CAPECCHI, M. R., EVANS, M. J. & SMITHIES, O. 2007. *The 2007 Nobel Prize in Physiology or Medicine - Advanced Information". Nobelprize.org. 15 Nov 2012* [Online]. Oslo. Available: http://www.nobelprize.org/nobel_prizes/medicine/laureates/2007/advanced.html.

CATANIA, E. H., PIMENTA, A. & LEVITT, P. 2008. Genetic deletion of Lsamp causes exaggerated behavioral activation in novel environments. *Behav Brain Res,* 188, 380-90.

CECIL, R. L., GOLDMAN, L. & AUSIELLO, D. A. 2004. *Cecil textbook of medicine,* Philadelphia, Saunders.

CHAPMAN, S. C., LAWSON, A., MACARTHUR, W. C., WIESE, R. J., LOECHEL, R. H., BURGOS-TRINIDAD, M., WAKEFIELD, J. K., RAMABHADRAN, R., MAUCH, T. J. & SCHOENWOLF, G. C. 2005. Ubiquitous GFP expression in transgenic chickens using a lentiviral vector. *Development,* 132, 935-40.

CHEN, J., LUI, W. O., VOS, M. D., CLARK, G. J., TAKAHASHI, M., SCHOUMANS, J., KHOO, S. K., PETILLO, D., LAVERY, T., SUGIMURA, J., ASTUTI, D., ZHANG, C., KAGAWA, S., MAHER, E. R., LARSSON, C., ALBERTS, A. S., KANAYAMA, H. O. & TEH, B. T. 2003. The t(1;3) breakpoint-spanning genes LSAMP and NORE1 are involved in clear cell renal cell carcinomas. *Cancer Cell,* 4, 405-13.

CHO, A., HARUYAMA, N. & KULKARNI, A. B. 2009. Generation of transgenic mice. *Curr Protoc Cell Biol,* Chapter 19, Unit 19 11.

CHURCH, D. M., GOODSTADT, L., HILLIER, L. W., ZODY, M. C., GOLDSTEIN, S., SHE, X., BULT, C. J., AGARWALA, R., CHERRY, J. L., DICUCCIO, M., HLAVINA, W., KAPUSTIN, Y., MERIC, P., MAGLOTT, D., BIRTLE, Z., MARQUES, A. C., GRAVES, T., ZHOU, S., TEAGUE, B., POTAMOUSIS, K., CHURAS, C., PLACE, M., HERSCHLEB, J., RUNNHEIM, R., FORREST, D., AMOS-LANDGRAF, J., SCHWARTZ, D. C., CHENG, Z., LINDBLAD-TOH, K., EICHLER, E. E., PONTING, C. P. & MOUSE GENOME SEQUENCING, C. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol,* 7, e1000112.

COLD SPRING HARBOR PROTOCOLS 2006. *Cold Spring Harbor Protocols [electronic journal],* [s.l.] : Cold Spring Harbor Laboratory Press.

CONSORTIUM, T. U. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research,* 40, D71-D75.

COOK, M. I., BEISSINGER, S. R., TORANZOS, G. A., RODRIGUEZ, R. A. & ARENDT, W. J. 2003. Trans-shell infection by pathogenic micro-organisms reduces the shelf life of non-incubated bird's eggs: a constraint on the onset of incubation? *Proceedings. Biological sciences / The Royal Society,* 270, 2233-40.

CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227, 561-3.

CRICK, F. H. 1958. On protein synthesis. *Symp Soc Exp Biol,* 12, 138-63.

CROTEAU, L. P. & KANIA, A. 2011. Optimisation of in ovo electroporation of the chick neural tube. *Journal of neuroscience methods,* 201, 381-4.

DAHL, U., SJODIN, A. & SEMB, H. 1996. Cadherins regulate aggregation of pancreatic beta-cells in vivo. *Development,* 122, 2895-902.

DARLING, S. M., CRAMPTON, J. M. & WILLIAMSON, R. 1982. Organization of a family of highly repetitive sequences within the human genome. *J Mol Biol,* 154, 51-63.

DEBOER, E., ANTONIOU, M., MIGNOTTE, V., WALL, L. & GROSVELD, F. 1988. The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *The EMBO journal,* 7, 4203-12.

DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science,* 295, 1306-11.

DHALLUIN, C., CARLSON, J. E., ZENG, L., HE, C., AGGARWAL, A. K. & ZHOU, M.-M. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature,* 399, 491-496.

DODGSON, J. B. 2003. Chicken genome sequence: a centennial gift to poultry genetics. *Cytogenet Genome Res,* 102, 291-6.

DOMSCHKE, K., DECKERT, J., O'DONOVAN M, C. & GLATT, S. J. 2007. Meta-analysis of COMT val158met in panic disorder: ethnic heterogeneity and gender specificity. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics,* 144B, 667-73.

DOOLITTLE, W. F. 1999. Phylogenetic Classification and the Universal Tree. *Science,* 284, 2124-2128.

ELKS, C. E., LOOS, R. J., SHARP, S. J., LANGENBERG, C., RING, S. M., TIMPSON, N. J., NESS, A. R., DAVEY SMITH, G., DUNGER, D. B., WAREHAM, N. J. & ONG, K. K. 2010. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med,* 7, e1000284.

EPPIG, J. T., BLAKE, J. A., BULT, C. J., KADIN, J. A., RICHARDSON, J. E. & MOUSE GENOME DATABASE, G. 2012. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res,* 40, D881-6.

ESCOFFRE, J. M., PORTET, T., WASUNGU, L., TEISSIE, J., DEAN, D. & ROLS, M. P. 2009. What is (still not) known of the mechanism by which electroporation mediates gene transfer and expression in cells and tissues. *Mol Biotechnol,* 41, 286-95.

ESNI, F., TALJEDAL, I. B., PERL, A. K., CREMER, H., CHRISTOFORI, G. & SEMB, H. 1999. Neural cell adhesion molecule (N-CAM) is required for cell type segregation and normal ultrastructure in pancreatic islets. *The Journal of cell biology,* 144, 325-37.

FEDERSPIEL, M. J. & HUGHES, S. H. 1997. Retroviral gene delivery. *Methods in cell biology,* 52, 179-214.

FERNÁNDEZ, M. & MIRANDA-SAAVEDRA, D. 2012. Genome-wide enhancer prediction from epigenetic signatures using genetic algorithm-optimized support vector machines. *Nucleic Acids Research.*

FLICEK, P., AMODE, M. R., BARRELL, D., BEAL, K., BRENT, S., CHEN, Y., CLAPHAM, P., COATES, G., FAIRLEY, S., FITZGERALD, S., GORDON, L., HENDRIX, M., HOURLIER, T., JOHNSON, N., KÄHÄRI, A., KEEFE, D., KEENAN, S., KINSELLA, R., KOKOCINSKI, F., KULESHA, E., LARSSON, P., LONGDEN, I., MCLAREN, W., OVERDUIN, B., PRITCHARD, B., RIAT, H. S., RIOS, D., RITCHIE, G. R. S., RUFFIER, M., SCHUSTER, M., SOBRAL, D., SPUDICH, G.,

TANG, Y. A., TREVANION, S., VANDROVCOVA, J., VILELLA, A. J., WHITE, S., WILDER, S. P., ZADISSA, A., ZAMORA, J., AKEN, B. L., BIRNEY, E., CUNNINGHAM, F., DUNHAM, I., DURBIN, R., FERNÁNDEZ-SUAREZ, X. M., HERRERO, J., HUBBARD, T. J. P., PARKER, A., PROCTOR, G., VOGEL, J. & SEARLE, S. M. J. 2011. Ensembl 2011. *Nucleic Acids Research,* 39, D800-D806.

FONDON, J. W., 3RD & GARNER, H. R. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A,* 101, 18058-63.

FORTES, P. & RAZQUIN, N. 2009. http://www.promega.com/resources/articles/pubhub/in-vivo-evaluation-of-regulatory-sequences-by-analysis-of-luciferase-expression/

FREEMAN, M. P., FREEMAN, S. A. & MCELROY, S. L. 2002. The comorbidity of bipolar and anxiety disorders: prevalence, psychobiology, and treatment issues. *Journal of affective disorders,* 68, 1-23.

FUJITA, H., MACHINO, M. & NAGATA, S. 1962. Electron microscopic studies on the development of the thyroid gland of the chick embryo. *Nihon Naibunpi Gakkai Zasshi,* 38, 725-9.

FUJITA, S. 1962. Kinetics of cellular proliferation. *Exp Cell Res,* 28, 52-60.

FUNATSU, N., MIYATA, S., KUMANOGOH, H., SHIGETA, M., HAMADA, K., ENDO, Y., SOKAWA, Y. & MAEKAWA, S. 1999. Characterization of a novel rat brain glycosylphosphatidylinositol-anchored protein (Kilon), a member of the IgLON cell adhesion molecule family. *J Biol Chem,* 274, 8224-30.

GABRIEL, B. & TEISSIE, J. 1997. Direct observation in the millisecond time range of fluorescent molecule asymmetrical interaction with the electropermeabilized cell membrane. *Biophys J,* 73, 2630-7.

GEYER, P. K., GREEN, M. M. & CORCES, V. G. 1990. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *The EMBO journal,* 9, 2247-56.

GIL, O. D., ZANAZZI, G., STRUYK, A. F. & SALZER, J. L. 1998. Neurotrimin mediates bifunctional effects on neurite outgrowth via homophilic and heterophilic interactions. *J Neurosci,* 18, 9312-25.

GOODMAN, C. S. & SHATZ, C. J. 1993. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell,* 72 Suppl, 77-98.

GOSSEN, M. & BUJARD, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A,* 89, 5547-51.

GOSSEN, M. & BUJARD, H. 2002. Studying gene function in eukaryotes by conditional gene inactivation. *Annu Rev Genet,* 36, 153-73.

GOSSEN, M., FREUNDLIEB, S., BENDER, G., MULLER, G., HILLEN, W. & BUJARD, H. 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science,* 268, 1766-9.

GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research,* 38, W695-9.

HACHISUKA, A., NAKAJIMA, O., YAMAZAKI, T. & SAWADA, J. 2000. Developmental expression of opioid-binding cell adhesion molecule (OBCAM) in rat brain. *Brain Res Dev Brain Res,* 122, 183-91.

HADCHOUEL, J., TAJBAKHSH, S., PRIMIG, M., CHANG, T. H., DAUBAS, P., ROCANCOURT, D. & BUCKINGHAM, M. 2000. Modular long-range regulation of Myf5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. *Development,* 127, 4455-67.
HALL, N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol,* 210, 1518-25.
HANCOX, K. A., GOOLEY, A. A. & JEFFREY, P. L. 1997. AvGp50, a predominantly axonally expressed glycoprotein, is a member of the IgLON's subfamily of cell adhesion molecules (CAMs). *Brain Res Mol Brain Res,* 44, 273-85.
HARADA, H., SUZU, S., HAYASHI, Y. & OKADA, S. 2005. BT-IgSF, a novel immunoglobulin superfamily protein, functions as a cell adhesion molecule. *J Cell Physiol,* 204, 919-26.
HARPAZ, Y. & CHOTHIA, C. 1994. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol,* 238, 528-39.
HARRIS, H. 1969. Enzyme and protein polymorphism in human populations. *Br Med Bull,* 25, 5-13.
HASHIMOTO, T., MAEKAWA, S. & MIYATA, S. 2009. IgLON cell adhesion molecules regulate synaptogenesis in hippocampal neurons. *Cell biochemistry and function,* 27, 496-8.
HASSEL, B., RATHJEN, F. G. & VOLKMER, H. 1997. Organization of the neurofascin gene and analysis of developmentally regulated alternative splicing. *J Biol Chem,* 272, 28742-9.
HAUPT, A., THAMER, C., HENI, M., MACHICAO, F., MACHANN, J., SCHICK, F., STEFAN, N., FRITSCHE, A., HARING, H. U. & STAIGER, H. 2010. Novel obesity risk loci do not determine distribution of body fat depots: a whole-body MRI/MRS study. *Obesity (Silver Spring),* 18, 1212-7.
HAY, C. W. & DOCHERTY, K. 2006. Comparative analysis of insulin gene promoters: implications for diabetes research. *Diabetes,* 55, 3201-13.
HEBENSTREIT, D., GIAISI, M., TREIBER, M. K., ZHANG, X. B., MI, H. F., HOREJS-HOECK, J., ANDERSEN, K. G., KRAMMER, P. H., DUSCHL, A. & LI-WEBER, M. 2008. LEF-1 negatively controls interleukin-4 expression through a proximal promoter regulatory element. *The Journal of biological chemistry,* 283, 22490-7.
HEDRICH, H. J. 2012. *The laboratory mouse,* Amsterdam ; London, Elsevier Academic Press.
HERMANN, B. P., HORNBAKER, K. I., MARAN, R. R. M. & HECKERT, L. L. 2007. Distal regulatory elements are required for Fshr expression, in vivo. *Molecular and Cellular Endocrinology,* 260–262, 49-58.
HIBINO, M., ITOH, H. & KINOSITA, K., JR. 1993. Time courses of cell electroporation as revealed by submicrosecond imaging of transmembrane potential. *Biophys J,* 64, 1789-800.
HIRANO, N., MUROI, T., TAKAHASHI, H. & HARUKI, M. 2011. Site-specific recombinases as tools for heterologous gene integration. *Applied microbiology and biotechnology,* 92, 227-39.
HODGKINSON, A. & EYRE-WALKER, A. 2011. Variation in the mutation rate across mammalian genomes. *Nature reviews. Genetics,* 12, 756-66.

HORTON, H. L. & LEVITT, P. 1988. A unique membrane protein is expressed on early developing limbic system axons and cortical targets. 8, 4653-61.
INADA, Y., YONEDA, H., KOH, J., SAKAI, J., HIMEI, A., KINOSHITA, Y., AKABAME, K., HIRAOKA, Y. & SAKAI, T. 2003. Positive association between panic disorder and polymorphism of the serotonin 2A receptor gene. *Psychiatry research,* 118, 25-31.
INNOS, J., PHILIPS, M. A., LEIDMAA, E., HEINLA, I., RAUD, S., REEMANN, P., PLAAS, M., NURK, K., KURRIKOFF, K., MATTO, V., VISNAPUU, T., MARDI, P., KOKS, S. & VASAR, E. 2011. Lower anxiety and a decrease in agonistic behaviour in Lsamp-deficient mice. *Behavioural brain research,* 217, 21-31.
INNOS, J., PHILIPS, M. A., RAUD, S., LILLEVALI, K., KOKS, S. & VASAR, E. 2012. Deletion of the Lsamp gene lowers sensitivity to stressful environmental manipulations in mice. *Behavioural brain research,* 228, 74-81.
INTERNATIONAL CHICKEN GENOME SEQUENCING, C. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432, 695-716.
ISLAM, M. M., DOH, S. T. & CAI, L. 2012. In ovo electroporation in embryonic chick retina. *Journal of visualized experiments : JoVE.*
ITASAKI, N., BEL-VIALAR, S. & KRUMLAUF, R. 1999. /`Shocking/' developments in chick embryology: electroporation and in ovo gene expression. *Nat Cell Biol,* 1, E203-E207.
IYER, A. K., MILLER, N. L., YIP, K., TRAN, B. H. & MELLON, P. L. 2010. Enhancers of GnRH transcription embedded in an upstream gene use homeodomain proteins to specify hypothalamic expression. *Mol Endocrinol,* 24, 1949-64.
JEFFREYS, A. J., WILSON, V. & THEIN, S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature,* 314, 67-73.
JOHNSON, C. P., FUJIMOTO, I., RUTISHAUSER, U. & LECKBAND, D. E. 2005. Direct evidence that neural cell adhesion molecule (NCAM) polysialylation increases intermembrane repulsion and abrogates adhesion. *The Journal of biological chemistry,* 280, 137-45.
JOHNSON, M., ZARETSKAYA, I., RAYTSELIS, Y., MEREZHUK, Y., MCGINNIS, S. & MADDEN, T. L. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res,* 36, W5-9.
JONTES, J. D. & SMITH, S. J. 2000. Filopodia, spines, and the generation of synaptic diversity. *Neuron,* 27, 11-4.
KALINOVSKY, A., BOUKHTOUCHE, F., BLAZESKI, R., BORNMANN, C., SUZUKI, N., MASON, C. A. & SCHEIFFELE, P. 2011. Development of Axon-Target Specificity of Ponto-Cerebellar Afferents. *PLoS Biol,* 9, e1001013.
KASPER, L. H., FUKUYAMA, T., BIESEN, M. A., BOUSSOUAR, F., TONG, C., DE PAUW, A., MURRAY, P. J., VAN DEURSEN, J. M. & BRINDLE, P. K. 2006. Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development. *Molecular and cellular biology,* 26, 789-809.
KATIDOU, M., VIDAKI, M., STRIGINI, M. & KARAGOGEOS, D. 2008. The immunoglobulin superfamily of neuronal cell adhesion molecules: lessons

from animal models and correlation with human disease. *Biotechnol J,* 3, 1564-80.
KATZ, L. C. & SHATZ, C. J. 1996. Synaptic activity and the construction of cortical circuits. *Science,* 274, 1133-8.
KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12, 996-1006.
KEREM, B., ROMMENS, J. M., BUCHANAN, J. A., MARKIEWICZ, D., COX, T. K., CHAKRAVARTI, A., BUCHWALD, M. & TSUI, L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science,* 245, 1073-80.
KESSLER, R. C., CHIU, W. T., DEMLER, O., MERIKANGAS, K. R. & WALTERS, E. E. 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry,* 62, 617-27.
KIMURA, Y., KATOH, A., KANEKO, T., TAKAHAMA, K. & TANAKA, H. 2001. Two members of the IgLON family are expressed in a restricted region of the developing chick brain and neural crest. *Dev Growth Differ,* 43, 257-63.
KING'S FUND, MCCRONE, P., DHANASIRI, S., PATEL, A., KNAPP, M. & LAWTON-SMITH, S. 2008. *Paying the Price - The cost of mental health care in England to 2026,* London.
KOENEN, K. C., AMSTADTER, A. B., RUGGIERO, K. J., ACIERNO, R., GALEA, S., KILPATRICK, D. G. & GELERNTER, J. 2009. RGS2 and generalized anxiety disorder in an epidemiologic sample of hurricane-exposed adults. *Depression and anxiety,* 26, 309-15.
KOIDO, K., KÕKS, S., MUST, A., REIMETS, A., MARON, E., SHLIK, J., VASAR, V. & VASAR, E. 2006. Association analysis of limbic system-associated membrane protein gene polymorphisms in mood and anxiety disorders. *European Neuropsychopharmacology,* 16, S9-S9.
KOIDO, K., TRAKS, T., BALOTSEV, R., ELLER, T., MUST, A., KOKS, S., MARON, E., TORU, I., SHLIK, J., VASAR, V. & VASAR, E. 2012. Associations between LSAMP gene polymorphisms and major depressive disorder and panic disorder. *Transl Psychiatry,* 2, e152.
KOO, B. C., KWON, M. S., CHOI, B. R., KIM, J. H., CHO, S. K., SOHN, S. H., CHO, E. J., LEE, H. T., CHANG, W., JEON, I., PARK, J. K., PARK, J. B. & KIM, T. 2006. Production of germline transgenic chickens expressing enhanced green fluorescent protein using a MoMLV-based retrovirus vector. *FASEB J,* 20, 2251-60.
KORBEL, J. O., URBAN, A. E., AFFOURTIT, J. P., GODWIN, B., GRUBERT, F., SIMONS, J. F., KIM, P. M., PALEJEV, D., CARRIERO, N. J., DU, L., TAILLON, B. E., CHEN, Z., TANZER, A., SAUNDERS, A. C., CHI, J., YANG, F., CARTER, N. P., HURLES, M. E., WEISSMAN, S. M., HARKINS, T. T., GERSTEIN, M. B., EGHOLM, M. & SNYDER, M. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science,* 318, 420-6.
KORESSAAR, T. & REMM, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics,* 23, 1289-91.
KOS, C. H. 2004. Cre/loxP system for generating tissue-specific knockout mouse models. *Nutrition reviews,* 62, 243-6.

KRESSE, S. H., OHNSTAD, H. O., PAULSEN, E. B., BJERKEHAGEN, B., SZUHAI, K., SERRA, M., SCHAEFER, K. L., MYKLEBOST, O. & MEZA-ZEPEDA, L. A. 2009. LSAMP, a novel candidate tumor suppressor gene in human osteosarcomas, identified by array comparative genomic hybridization. *Genes, chromosomes & cancer,* 48, 679-93.

KRIZSAN-AGBAS, D., PEDCHENKO, T. & SMITH, P. G. 2008. Neurotrimin is an estrogen-regulated determinant of peripheral sympathetic innervation. *Journal of neuroscience research,* 86, 3086-95.

KUHN, M. W., RADTKE, I., BULLINGER, L., GOORHA, S., CHENG, J., EDELMANN, J., GOHLKE, J., SU, X., PASCHKA, P., POUNDS, S., KRAUTER, J., GANSER, A., QUESSAR, A., RIBEIRO, R., GAIDZIK, V. I., SHURTLEFF, S., KRONKE, J., HOLZMANN, K., MA, J., SCHLENK, R. F., RUBNITZ, J. E., DOHNER, K., DOHNER, H. & DOWNING, J. R. 2012. High-resolution genomic profiling of adult and pediatric core-binding factor acute myeloid leukemia reveals new recurrent genomic alterations. *Blood,* 119, e67-75.

KUMAR, R., THOMAS A. FIRAK, CHRISTOPHER T. SCHROLL & KIRANUR N. SUBRAMANIAN 1986. Activation of gene expression is adversely affected at high multiplicities of linked simian virus 40 enhancer. *Proc. Natl. Acad. Sci. USA,* 83.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23, 2947-8.

LAWSON, P. T. 2000. *Assistant Laboratory Animal Technician Training Manual* American Association for Laboratory Animal Science.

LEE, K. J., DIETRICH, P. & JESSELL, T. M. 2000. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. *Nature,* 403, 734-40.

LEE, M. K., REN, C. W., YAN, B., COX, B., ZHANG, H. B., ROMANOV, M. N., SIZEMORE, F. G., SUCHYTA, S. P., PETERS, E. & DODGSON, J. B. 2003. Construction and characterization of three BAC libraries for analysis of the chicken genome. *Animal genetics,* 34, 151-2.

LESHCHYNS'KA, I., SYTNYK, V., MORROW, J. S. & SCHACHNER, M. 2003. Neural cell adhesion molecule (NCAM) association with PKCbeta2 via betaI spectrin is implicated in NCAM-mediated neurite outgrowth. *J Cell Biol,* 161, 625-39.

LESHCHYNS'KA, I., SYTNYK, V., RICHTER, M., ANDREYEVA, A., PUCHKOV, D. & SCHACHNER, M. 2006. The adhesion molecule CHL1 regulates uncoating of clathrin-coated synaptic vesicles. *Neuron,* 52, 1011-25.

LEVITT, P. 1984. A monoclonal antibody to limbic system neurons. *Science,* 223, 299-301.

LI, H., ARBER, S., JESSELL, T. M. & EDLUND, H. 1999. Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. *Nat Genet,* 23, 67-70.

LI, P., PRASAD, S. S., MITCHELL, D. E., HACHISUKA, A., SAWADA, J. I., AL-HOUSSEINI, A. M. & GU, Q. 2006. Postnatal expression profile of

OBCAM implies its involvement in visual cortex development and plasticity. *Cerebral cortex,* 16, 291-9.

LI, W. H. & WU, C. I. 1987. Rates of nucleotide substitution are evidently higher in rodents than in man. *Molecular biology and evolution,* 4, 74-82.

LINDBLAD-TOH, K., GARBER, M., ZUK, O., LIN, M. F., PARKER, B. J., WASHIETL, S., KHERADPOUR, P., ERNST, J., JORDAN, G., MAUCELI, E., WARD, L. D., LOWE, C. B., HOLLOWAY, A. K., CLAMP, M., GNERRE, S., ALFOLDI, J., BEAL, K., CHANG, J., CLAWSON, H., CUFF, J., DI PALMA, F., FITZGERALD, S., FLICEK, P., GUTTMAN, M., HUBISZ, M. J., JAFFE, D. B., JUNGREIS, I., KENT, W. J., KOSTKA, D., LARA, M., MARTINS, A. L., MASSINGHAM, T., MOLTKE, I., RANEY, B. J., RASMUSSEN, M. D., ROBINSON, J., STARK, A., VILELLA, A. J., WEN, J., XIE, X., ZODY, M. C., BROAD INSTITUTE SEQUENCING, P., WHOLE GENOME ASSEMBLY, T., BALDWIN, J., BLOOM, T., CHIN, C. W., HEIMAN, D., NICOL, R., NUSBAUM, C., YOUNG, S., WILKINSON, J., WORLEY, K. C., KOVAR, C. L., MUZNY, D. M., GIBBS, R. A., BAYLOR COLLEGE OF MEDICINE HUMAN GENOME SEQUENCING CENTER SEQUENCING, T., CREE, A., DIHN, H. H., FOWLER, G., JHANGIANI, S., JOSHI, V., LEE, S., LEWIS, L. R., NAZARETH, L. V., OKWUONU, G., SANTIBANEZ, J., WARREN, W. C., MARDIS, E. R., WEINSTOCK, G. M., WILSON, R. K., GENOME INSTITUTE AT WASHINGTON, U., DELEHAUNTY, K., DOOLING, D., FRONIK, C., FULTON, L., FULTON, B., GRAVES, T., MINX, P., SODERGREN, E., BIRNEY, E., MARGULIES, E. H., HERRERO, J., GREEN, E. D., HAUSSLER, D., SIEPEL, A., GOLDMAN, N., POLLARD, K. S., PEDERSEN, J. S., LANDER, E. S. & KELLIS, M. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature,* 478, 476-82.

LIU, H., FOCIA, P. J. & HE, X. 2011. Homophilic adhesion mechanism of neurofascin, a member of the L1 family of neural cell adhesion molecules. *J Biol Chem,* 286, 797-805.

LODGE, A. P., MCNAMEE, C. J., HOWARD, M. R., REED, J. E. & MOSS, D. J. 2001a. Identification and characterization of CEPU-Se-A secreted isoform of the IgLON family protein, CEPU-1. *Mol Cell Neurosci,* 17, 746-60.

LODGE, A. P., MCNAMEE, C. J., HOWARD, M. R., REED, J. E. & MOSS, D. J. 2001b. Identification and characterization of CEPU-Se-A secreted isoform of the IgLON family protein, CEPU-1. *Molecular and cellular neurosciences,* 17, 746-60.

LOMVARDAS, S., BARNEA, G., PISAPIA, D. J., MENDELSOHN, M., KIRKLAND, J. & AXEL, R. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell,* 126, 403-13.

LONGO, D. L. & HARRISON, T. R. 2012. *Harrison's principles of internal medicine,* New York, McGraw-Hill.

LOOTS, G. G. & OVCHARENKO, I. 2004. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res,* 32, W217-21.

LOPEZ, A., ROLS, M. P. & TEISSIE, J. 1988. 31P NMR analysis of membrane phospholipid organization in viable, reversibly electropermeabilized Chinese hamster ovary cells. *Biochemistry,* 27, 1222-8.

MACDONALD, M. E., AMBROSE, C. M., DUYAO, M. P., MYERS, R. H., LIN, C., SRINIDHI, L., BARNES, G., TAYLOR, S. A., JAMES, M., GROOT, N., MACFARLANE, H., JENKINS, B., ANDERSON, M. A., WEXLER, N. S., GUSELLA, J. F., BATES, G. P., BAXENDALE, S., HUMMERICH, H., KIRBY, S., NORTH, M., YOUNGMAN, S., MOTT, R., ZEHETNER, G., SEDLACEK, Z., POUSTKA, A., FRISCHAUF, A. M., LEHRACH, H., BUCKLER, A. J., CHURCH, D., DOUCETTESTAMM, L., ODONOVAN, M. C., RIBARAMIREZ, L., SHAH, M., STANTON, V. P., STROBEL, S. A., DRATHS, K. M., WALES, J. L., DERVAN, P., HOUSMAN, D. E., ALTHERR, M., SHIANG, R., THOMPSON, L., FIELDER, T., WASMUTH, J. J., TAGLE, D., VALDES, J., ELMER, L., ALLARD, M., CASTILLA, L., SWAROOP, M., BLANCHARD, K., COLLINS, F. S., SNELL, R., HOLLOWAY, T., GILLESPIE, K., DATSON, N., SHAW, D. & HARPER, P. S. 1993. A NOVEL GENE CONTAINING A TRINUCLEOTIDE REPEAT THAT IS EXPANDED AND UNSTABLE ON HUNTINGTONS-DISEASE CHROMOSOMES. *Cell,* 72, 971-983.

MAEKAWA, T., IMAMOTO, F., MERLINOT, G. T., PASTANG, I. & S, I. 1989. Cooperative Function of Two Separate Enhancers of the Human Epidermal Growth Factor Receptor Proto-oncoge. *THE JOURNAL OF BIOLOGICAL CHEMISTRY,* 10.

MAGRANE, M. & CONSORTIUM, U. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database,* 2011.

MALIK, S. & ROEDER, R. G. 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics,* 11, 761-72.

MANESS, P. F., BEGGS, H. E., KLINZ, S. G. & MORSE, W. R. 1996. Selective neural cell adhesion molecule signaling by Src family tyrosine kinases and tyrosine phosphatases. *Perspect Dev Neurobiol,* 4, 169-81.

MANESS, P. F. & SCHACHNER, M. 2007. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat Neurosci,* 10, 19-26.

MANN, F., ZHUKAREVA, V., PIMENTA, A., LEVITT, P. & BOLZ, J. 1998. Membrane-associated molecules guide limbic and nonlimbic thalamocortical projections. *J Neurosci,* 18, 9409-19.

MARG, A., SIRIM, P., SPALTMANN, F., PLAGGE, A., KAUSELMANN, G., BUCK, F., RATHJEN, F. G. & BRUMMENDORF, T. 1999. Neurotractin, a novel neurite outgrowth-promoting Ig-like protein that interacts with CEPU-1 and LAMP. *J Cell Biol,* 145, 865-76.

MARKS, I. M. & NESSE, R. M. 1994. Fear and Fitness: An Evolutionary Analysis of Anxiety Disorders. *Ethology and Sociobiology* 15.

MARON, E., KOIDO, K., MUST, A., REIMETS, A., KÖKS, S., VASAR, E., VASAR, V. & SHLIK, J. 2006. P.4.a.022 Association study of limbic system-associated membrane protein gene polymorphisms in panic disorder. *European Neuropsychopharmacology,* 16, S459-S460.

MARON, E., NIKOPENSIUS, T., KOKS, S., ALTMAE, S., HEINASTE, E., VABRIT, K., TAMMEKIVI, V., HALLAST, P., KOIDO, K., KURG, A., METSPALU, A., VASAR, E., VASAR, V. & SHLIK, J. 2005. Association study of 90 candidate gene polymorphisms in panic disorder. *Psychiatric genetics,* 15, 17-24.

MASTON, G. A., EVANS, S. K. & GREEN, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics,* 7, 29-59.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34, D108-10.

MCGREW, M. J., SHERMAN, A., ELLARD, F. M., LILLICO, S. G., GILHOOLEY, H. J., KINGSMAN, A. J., MITROPHANOUS, K. A. & SANG, H. 2004. Efficient production of germline transgenic chickens using lentiviral vectors. *EMBO reports,* 5, 728-33.

MCNAMEE, C. J., REED, J. E., HOWARD, M. R., LODGE, A. P. & MOSS, D. J. 2002. Promotion of neuronal cell adhesion by members of the IgLON family occurs in the absence of either support or modification of neurite outgrowth. *J Neurochem,* 80, 941-8.

MCNAMEE, C. J., YOUSSEF, S. & MOSS, D. 2011. IgLONs form heterodimeric complexes on forebrain neurons. *Cell biochemistry and function,* 29, 114-9.

MENDELSON, S. C. & QUINN, J. P. 1993. Identification of potential regulatory elements within the rat preprotachykinin A promoter. *Biochem Soc Trans,* 21, 372S.

METZENBERG, S. 2007. *Working with DNA,* New York, Taylor & Francis Group.

MILEWSKI, R. C., CHI, N. C., LI, J., BROWN, C., LU, M. M. & EPSTEIN, J. A. 2004. Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of Tead2 as a regulator of Pax3. *Development,* 131, 829-837.

MILLER, E. M. & NICKOLOFF, J. A. 1995. Escherichia coli electrotransformation. *Methods Mol Biol,* 47, 105-13.

MIMS. 2012. *Buspirone* [Online]. Haymarket. Available: http://www.mims.co.uk/Drugs/central-nervous-system/anxiety/buspirone/.

MIYATA, S., MATSUMOTO, N., TAGUCHI, K., AKAGI, A., IINO, T., FUNATSU, N. & MAEKAWA, S. 2003a. Biochemical and ultrastructural analyses of IgLON cell adhesion molecules, Kilon and OBCAM in the rat brain. *Neuroscience,* 117, 645-58.

MIYATA, S., TAGUCHI, K. & MAEKAWA, S. 2003b. Dendrite-associated opioid-binding cell adhesion molecule localizes at neurosecretory granules in the hypothalamic magnocellular neurons. *Neuroscience,* 122, 169-81.

MOMOSE, T., TONEGAWA, A., TAKEUCHI, J., OGAWA, H., UMESONO, K. & YASUDA, K. 1999. Efficient targeting of gene expression in chick embryos by microelectroporation. *Dev Growth Differ,* 41, 335-44.

MONTAG-SALLAZ, M., BAARKE, A. & MONTAG, D. 2003. Aberrant neuronal connectivity in CHL1-deficient mice is associated with altered information processing-related immediate early gene expression. *J Neurobiol,* 57, 67-80.

MOUSE GENOME SEQUENCING, C., WATERSTON, R. H., LINDBLAD-TOH, K., BIRNEY, E., ROGERS, J., ABRIL, J. F., AGARWAL, P., AGARWALA, R., AINSCOUGH, R., ALEXANDERSSON, M., AN, P.,

ANTONARAKIS, S. E., ATTWOOD, J., BAERTSCH, R., BAILEY, J., BARLOW, K., BECK, S., BERRY, E., BIRREN, B., BLOOM, T., BORK, P., BOTCHERBY, M., BRAY, N., BRENT, M. R., BROWN, D. G., BROWN, S. D., BULT, C., BURTON, J., BUTLER, J., CAMPBELL, R. D., CARNINCI, P., CAWLEY, S., CHIAROMONTE, F., CHINWALLA, A. T., CHURCH, D. M., CLAMP, M., CLEE, C., COLLINS, F. S., COOK, L. L., COPLEY, R. R., COULSON, A., COURONNE, O., CUFF, J., CURWEN, V., CUTTS, T., DALY, M., DAVID, R., DAVIES, J., DELEHAUNTY, K. D., DERI, J., DERMITZAKIS, E. T., DEWEY, C., DICKENS, N. J., DIEKHANS, M., DODGE, S., DUBCHAK, I., DUNN, D. M., EDDY, S. R., ELNITSKI, L., EMES, R. D., ESWARA, P., EYRAS, E., FELSENFELD, A., FEWELL, G. A., FLICEK, P., FOLEY, K., FRANKEL, W. N., FULTON, L. A., FULTON, R. S., FUREY, T. S., GAGE, D., GIBBS, R. A., GLUSMAN, G., GNERRE, S., GOLDMAN, N., GOODSTADT, L., GRAFHAM, D., GRAVES, T. A., GREEN, E. D., GREGORY, S., GUIGO, R., GUYER, M., HARDISON, R. C., HAUSSLER, D., HAYASHIZAKI, Y., HILLIER, L. W., HINRICHS, A., HLAVINA, W., HOLZER, T., HSU, F., HUA, A., HUBBARD, T., HUNT, A., JACKSON, I., JAFFE, D. B., JOHNSON, L. S., JONES, M., JONES, T. A., JOY, A., KAMAL, M., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature,* 420, 520-62.

MOZDZIAK, P. E., BORWORNPINYO, S., MCCOY, D. W. & PETITTE, J. N. 2003. Development of transgenic chickens expressing bacterial beta-galactosidase. *Dev Dyn,* 226, 439-45.

MURAMATSU, T., MIZUTANI, Y., OHMORI, Y. & OKUMURA, J. 1997. Comparison of three nonviral transfection methods for foreign gene expression in early chicken embryos in ovo. *Biochemical and biophysical research communications,* 230, 376-80.

MUST, A., TASA, G., LANG, A., VASAR, E., KOKS, S., MARON, E. & VALI, M. 2008. Association of limbic system-associated membrane protein (LSAMP) to male completed suicide. *BMC Med Genet,* 9, 34.

MUZNY, D. M., SCHERER, S. E., KAUL, R., WANG, J., YU, J., SUDBRAK, R., BUHAY, C. J., CHEN, R., CREE, A., DING, Y., DUGAN-ROCHA, S., GILL, R., GUNARATNE, P., HARRIS, R. A., HAWES, A. C., HERNANDEZ, J., HODGSON, A. V., HUME, J., JACKSON, A., KHAN, Z. M., KOVAR-SMITH, C., LEWIS, L. R., LOZADO, R. J., METZKER, M. L., MILOSAVLJEVIC, A., MINER, G. R., MORGAN, M. B., NAZARETH, L. V., SCOTT, G., SODERGREN, E., SONG, X. Z., STEFFEN, D., WEI, S., WHEELER, D. A., WRIGHT, M. W., WORLEY, K. C., YUAN, Y., ZHANG, Z., ADAMS, C. Q., ANSARI-LARI, M. A., AYELE, M., BROWN, M. J., CHEN, G., CHEN, Z., CLENDENNING, J., CLERC-BLANKENBURG, K. P., DAVIS, C., DELGADO, O., DINH, H. H., DONG, W., DRAPER, H., ERNST, S., FU, G., GONZALEZ-GARAY, M. L., GARCIA, D. K., GILLETT, W., GU, J., HAO, B., HAUGEN, E., HAVLAK, P., HE, X., HENNIG, S., HU, S., HUANG, W., JACKSON, L. R., JACOB, L. S., KELLY, S. H., KUBE, M., LEVY, R., LI, Z., LIU, B., LIU, J., LIU, W., LU, J., MAHESHWARI, M., NGUYEN, B. V., OKWUONU, G. O., PALMEIRI, A., PASTERNAK, S., PEREZ, L. M., PHELPS, K. A., PLOPPER, F. J., QIANG, B., RAYMOND, C., RODRIGUEZ, R., SAENPHIMMACHAK, C., SANTIBANEZ, J., SHEN,

H., SHEN, Y., SUBRAMANIAN, S., TABOR, P. E., VERDUZCO, D., WALDRON, L., WANG, Q., WILLIAMS, G. A., WONG, G. K., YAO, Z., ZHANG, J., ZHANG, X., ZHAO, G., et al. 2006. The DNA sequence, annotation and analysis of human chromosome 3. *Nature,* 440, 1194-8.

NAGY, A. 2003. *Manipulating the mouse embryo : a laboratory manual,* Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

NAGY, A., GERTSENSTEIN, M., VINTERSTEN, K. & BEHRINGER, R. 2007. Staining Mouse Embryos for Alkaline Phosphatase Activity. *Cold Spring Harbor Protocols,* 2007, pdb.prot4776-.

NAKAJIMA, O., HACHISUKA, A., TAKAGI, K., YAMAZAKI, T., IKEBUCHI, H. & SAWADA, J. 1997. Expression of opioid-binding cell adhesion molecule (OBCAM) and neurotrimin (NTM) in E. coli and their reactivity with monoclonal anti-OBCAM antibody. *Neuroreport,* 8, 3005-8.

NAKAMURA, Y. 2009. DNA variations in human and medical genetics: 25 years of my experience. *J Hum Genet,* 54, 1-8.

NAKANO, T., WINDREM, M., ZAPPAVIGNA, V. & GOLDMAN, S. A. 2005. Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons. *Dev Biol,* 283, 474-85.

NCBI-HOMOLOGENE 2011. HomoloGene.

NELOVKOV, A., AREDA, T., INNOS, J., KOKS, S. & VASAR, E. 2006. Rats displaying distinct exploratory activity also have different expression patterns of gamma-aminobutyric acid- and cholecystokinin-related genes in brain regions. *Brain Res,* 1100, 21-31.

NELOVKOV, A., PHILIPS, M. A., KOKS, S. & VASAR, E. 2003. Rats with low exploratory activity in the elevated plus-maze have the increased expression of limbic system-associated membrane protein gene in the periaqueductal grey. *Neurosci Lett,* 352, 179-82.

NIEUWENHUYS, R., DONKELAAR, H. J. T. & NICHOLSON, C. 1998. *The central nervous system of vertebrates,* New York, Springer.

NORTHCUTT, R. G. 2002. Understanding vertebrate brain evolution. *Integr Comp Biol,* 42, 743-56.

NTOUGKOS, E., RUSH, R., SCOTT, D., FRANKENBERG, T., GABRA, H., SMYTH, J. F. & SELLAR, G. C. 2005. The IgLON family in epithelial ovarian cancer: expression profiles and clinicopathologic correlates. *Clin Cancer Res,* 11, 5764-8.

OKADA, F., SASA, H., OTANI, K., KANEKO, S., FUKUSHIMA, Y. & NOMURA, K. 1993. Letters. *Human Psychopharmacology: Clinical and Experimental,* 8, 371-372.

OLSSON, A. & PHELPS, E. A. 2007. Social learning of fear. *Nat Neurosci,* 10, 1095-102.

ORHAN, I. O. & ULUŞAHIN, A. 2001. Depression, anxiety comorbidity, and disability in tuberculosis and chronic obstructive pulmonary disease patients: applicability of GHQ-12. *General hospital psychiatry,* 23, 77-83.

OSBORN, L., VASSALLO, C., BROWNING, B. G., TIZARD, R., HASKARD, D. O., BENJAMIN, C. D., DOUGAS, I. & KIRCHHAUSEN, T. 1994. Arrangement of domains, and amino acid residues required for binding of vascular cell adhesion molecule-1 to its counter-receptor VLA-4 (alpha 4 beta 1). *The Journal of cell biology,* 124, 601-8.

OVCHARENKO, I., NOBREGA, M. A., LOOTS, G. G. & STUBBS, L. 2004. ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res,* 32, W280-6.
PAN, Y., WANG, K. S. & ARAGAM, N. 2011. NTM and NR3C2 polymorphisms influencing intelligence: family-based association studies. *Progress in neuro-psychopharmacology & biological psychiatry,* 35, 154-60.
PAPADAKIS, M. A., MCPHEE, S. J. & RABOW, M. W. 2013. Current medical diagnosis & treatment 2013. 52nd ed. New York: McGraw-Hill Medical.
PARATCHA, G., LEDDA, F. & IBANEZ, C. F. 2003. The neural cell adhesion molecule NCAM is an alternative signaling receptor for GDNF family ligands. *Cell,* 113, 867-79.
PASIC, I., SHLIEN, A., DURBIN, A. D., STAVROPOULOS, D. J., BASKIN, B., RAY, P. N., NOVOKMET, A. & MALKIN, D. 2010. Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res,* 70, 160-71.
PATZKE, C., MAX, K. E., BEHLKE, J., SCHREIBER, J., SCHMIDT, H., DORNER, A. A., KROGER, S., HENNING, M., OTTO, A., HEINEMANN, U. & RATHJEN, F. G. 2010. The coxsackievirus-adenovirus receptor reveals complex homophilic and heterophilic interactions on neural cells. *J Neurosci,* 30, 2897-910.
PDRAW32 BY ACACLONE SOFTWARE. Available: (http://www.acaclone.com).
PEARTON, D. J., BROADHURST, R., DONNISON, M. & PFEFFER, P. L. 2011. Elf5 regulation in the trophectoderm. *Dev Biol,* 360, 343-50.
PENNACCHIO, L. A., AHITUV, N., MOSES, A. M., PRABHAKAR, S., NOBREGA, M. A., SHOUKRY, M., MINOVITSKY, S., DUBCHAK, I., HOLT, A., LEWIS, K. D., PLAJZER-FRICK, I., AKIYAMA, J., DE VAL, S., AFZAL, V., BLACK, B. L., COURONNE, O., EISEN, M. B., VISEL, A. & RUBIN, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature,* 444, 499-502.
PEREZ-PEREZ, J. M., CANDELA, H. & MICOL, J. L. 2009. Understanding synergy in genetic interactions. *Trends Genet,* 25, 368-76.
PIMENTA, A. F., FISCHER, I. & LEVITT, P. 1996. cDNA cloning and structural analysis of the human limbic-system-associated membrane protein (LAMP). *Gene,* 170, 189-95.
PIMENTA, A. F. & LEVITT, P. 2004. Characterization of the genomic structure of the mouse limbic system-associated membrane protein (Lsamp) gene. *Genomics,* 83, 790-801.
PIMENTA, A. F., ZHUKAREVA, V., BARBE, M. F., REINOSO, B. S., GRIMLEY, C., HENZEL, W., FISCHER, I. & LEVITT, P. 1995. The limbic system-associated membrane protein is an Ig superfamily member that mediates selective neuronal growth and axon targeting. *Neuron,* 15, 287-97.
PORCU, S., KITAMURA, M., WITKOWSKA, E., ZHANG, Z., MUTERO, A., LIN, C., CHANG, J. & GAENSLER, K. M. 1997. The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood,* 90, 4602-9.
PRABHAKAR, S., POULIN, F., SHOUKRY, M., AFZAL, V., RUBIN, E. M., COURONNE, O. & PENNACCHIO, L. A. 2006. Close sequence

comparisons are sufficient to identify human cis-regulatory elements. *Genome Res,* 16, 855-63.

PRASAD, T. S., KANDASAMY, K. & PANDEY, A. 2009. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods in molecular biology,* 577, 67-79.

PUCIHAR, G., KOTNIK, T., VALIC, B. & MIKLAVCIC, D. 2006. Numerical determination of transmembrane voltage induced on irregularly shaped cells. *Annals of biomedical engineering,* 34, 642-52.

QIU, S., CHAMPAGNE, D. L., PETERS, M., CATANIA, E. H., WEEBER, E. J., LEVITT, P. & PIMENTA, A. F. 2010. Loss of limbic system-associated membrane protein leads to reduced hippocampal mineralocorticoid receptor expression, impaired synaptic plasticity, and spatial memory deficit. *Biological psychiatry,* 68, 197-204.

RABQUER, B. J., AMIN, M. A., TEEGALA, N., SHAHEEN, M. K., TSOU, P. S., RUTH, J. H., LESCH, C. A., IMHOF, B. A. & KOCH, A. E. 2010. Junctional adhesion molecule-C is a soluble mediator of angiogenesis. *J Immunol,* 185, 1777-85.

RANSCHT, B. & DOURS-ZIMMERMANN, M. T. 1991. T-cadherin, a novel cadherin cell adhesion molecule in the nervous system lacks the conserved cytoplasmic region. *Neuron,* 7, 391-402.

RAVASI, T., SUZUKI, H., CANNISTRACI, C. V., KATAYAMA, S., BAJIC, V. B., TAN, K., AKALIN, A., SCHMEIER, S., KANAMORI-KATAYAMA, M., BERTIN, N., CARNINCI, P., DAUB, C. O., FORREST, A. R., GOUGH, J., GRIMMOND, S., HAN, J. H., HASHIMOTO, T., HIDE, W., HOFMANN, O., KAMBUROV, A., KAUR, M., KAWAJI, H., KUBOSAKI, A., LASSMANN, T., VAN NIMWEGEN, E., MACPHERSON, C. R., OGAWA, C., RADOVANOVIC, A., SCHWARTZ, A., TEASDALE, R. D., TEGNER, J., LENHARD, B., TEICHMANN, S. A., ARAKAWA, T., NINOMIYA, N., MURAKAMI, K., TAGAMI, M., FUKUDA, S., IMAMURA, K., KAI, C., ISHIHARA, R., KITAZUME, Y., KAWAI, J., HUME, D. A., IDEKER, T. & HAYASHIZAKI, Y. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell,* 140, 744-52.

RAVINA, B., ROMER, M., CONSTANTINESCU, R., BIGLAN, K., BROCHT, A., KIEBURTZ, K., SHOULSON, I. & MCDERMOTT, M. P. 2008. The relationship between CAG repeat length and clinical progression in Huntington's disease. *Movement Disorders,* 23, 1223-1227.

RAY, P. & GAMBHIR, S. S. 2007. Noninvasive imaging of molecular events with bioluminescent reporter genes in living subjects. *Methods in molecular biology,* 411, 131-44.

REED, J., MCNAMEE, C., RACKSTRAW, S., JENKINS, J. & MOSS, D. 2004. Diglons are heterodimeric proteins composed of IgLON subunits, and Diglon-CO inhibits neurite outgrowth from cerebellar granule cells. *J Cell Sci,* 117, 3961-73.

REED, J. E., DUNN, J. R., DU PLESSIS, D. G., SHAW, E. J., REEVES, P., GEE, A. L., WARNKE, P. C., SELLAR, G. C., MOSS, D. J. & WALKER, C. 2007. Expression of cellular adhesion molecule 'OPCML' is down-regulated in gliomas and other brain tumours. *Neuropathol Appl Neurobiol,* 33, 77-85.

REINOSO, B. S., PIMENTA, A. F. & LEVITT, P. 1996. Expression of the mRNAs encoding the limbic system-associated membrane protein (LAMP): I. Adult rat brain. *J Comp Neurol,* 375, 274-88.

RENSTROM, F., PAYNE, F., NORDSTROM, A., BRITO, E. C., ROLANDSSON, O., HALLMANS, G., BARROSO, I., NORDSTROM, P. & FRANKS, P. W. 2009. Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Hum Mol Genet,* 18, 1489-96.

REVEST, J. M., FAIVRE-SARRAILH, C., MAEDA, N., NODA, M., SCHACHNER, M. & ROUGON, G. 1999. The interaction between F3 immunoglobulin domains and protein tyrosine phosphatases zeta/beta triggers bidirectional signalling between neurons and glial cells. *Eur J Neurosci,* 11, 1134-47.

RGD 2011. Rat Genome Database Web Site. Medical College of Wisconsin, Milwaukee, Wisconsin.

RIEDLE, S., KIEFEL, H., GAST, D., BONDONG, S., WOLTERINK, S., GUTWEIN, P. & ALTEVOGT, P. 2009. Nuclear translocation and signalling of L1-CAM in human carcinoma cells requires ADAM10 and presenilin/gamma-secretase activity. *Biochem J,* 420, 391-402.

RIORDAN, J. R., ROMMENS, J. M., KEREM, B., ALON, N., ROZMAHEL, R., GRZELCZAK, Z., ZIELENSKI, J., LOK, S., PLAVSIC, N., CHOU, J. L. & ET AL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science,* 245, 1066-73.

ROMMENS, J. M., IANNUZZI, M. C., KEREM, B., DRUMM, M. L., MELMER, G., DEAN, M., ROZMAHEL, R., COLE, J. L., KENNEDY, D., HIDAKA, N. & ET AL. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science,* 245, 1059-65.

ROSS, A. J., RUIZ-PEREZ, V., WANG, Y., HAGAN, D. M., SCHERER, S., LYNCH, S. A., LINDSAY, S., CUSTARD, E., BELLONI, E., WILSON, D. I., WADEY, R., GOODMAN, F., ORSTAVIK, K. H., MONCLAIR, T., ROBSON, S., REARDON, W., BURN, J., SCAMBLER, P. & STRACHAN, T. 1998. A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. *Nat Genet,* 20, 358-61.

ROUGON, G. & HOBERT, O. 2003. New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu Rev Neurosci,* 26, 207-38.

ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132, 365-86.

SADOUL, K., BOYAULT, C., PABION, M. & KHOCHBIN, S. 2008. Regulation of protein turnover by acetyltransferases and deacetylases. *Biochimie,* 90, 306-312.

SAINSBURY CENTRE FOR MENTAL HEALTH 2011. Annual Review - 2011. London: Sainsbury Centre for Mental Health.

SAKURAI, K., MIGITA, O., TORU, M. & ARINAMI, T. 2002. An association between a missense polymorphism in the close homologue of L1 (CHL1, CALL) gene and schizophrenia. *Mol Psychiatry,* 7, 412-5.

SALLEE, J. L., WITTCHEN, E. S. & BURRIDGE, K. 2006. Regulation of cell adhesion by protein-tyrosine phosphatases: II. Cell-cell adhesion. *J Biol Chem,* 281, 16189-92.

SAMBROOK, J., MANIATIS, T., FRITSCH, E. F. & LABORATORY, C. S. H. 1989. *Molecular cloning : a laboratory manual* Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 1989.

SAUER, B. 1987. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol,* 7, 2087-96.

SAUER, B. & HENDERSON, N. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A,* 85, 5166-70.

SCHAFER, M., BRAUER, A. U., SAVASKAN, N. E., RATHJEN, F. G. & BRUMMENDORF, T. 2005. Neurotractin/kilon promotes neurite outgrowth and is expressed on reactive astrocytes after entorhinal cortex lesion. *Mol Cell Neurosci,* 29, 580-90.

SCHLAKE, T. & BODE, J. 1994. Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry,* 33, 12746-51.

SCHLEINITZ, N., COGNET, C., GUIA, S., LAUGIER-ANFOSSI, F., BARATIN, M., POUGET, J., PELISSIER, J. F., HARLE, J. R., VIVIER, E. & FIGARELLA-BRANGER, D. 2008. Expression of the CD85j (leukocyte Ig-like receptor 1, Ig-like transcript 2) receptor for class I major histocompatibility complex molecules in idiopathic inflammatory myopathies. *Arthritis and rheumatism,* 58, 3216-23.

SCHMID, C. W. & DEININGER, P. L. 1975. Sequence organization of the human genome. *Cell,* 6, 345-58.

SCHMID, R. S. & MANESS, P. F. 2008. L1 and NCAM adhesion molecules as signaling coreceptors in neuronal migration and process outgrowth. *Current opinion in neurobiology,* 18, 245-250.

SCHNEIDER, B. F. & NORTON, S. 1979. Equivalent ages in rat, mouse and chick embryos. *Teratology,* 19, 273-8.

SCHOENFELDER, S., CLAY, I. & FRASER, P. 2010. The transcriptional interactome: gene expression in 3D. *Current opinion in genetics & development,* 20, 127-33.

SCHOFIELD, P. R., MCFARLAND, K. C., HAYFLICK, J. S., WILCOX, J. N., CHO, T. M., ROY, S., LEE, N. M., LOH, H. H. & SEEBURG, P. H. 1989. Molecular characterization of a new immunoglobulin superfamily protein with potential roles in opioid binding and cell contact. *Embo J,* 8, 489-95.

SCHUSTER, T., KRUG, M., STALDER, M., HACKEL, N., GERARDY-SCHAHN, R. & SCHACHNER, M. 2001. Immunoelectron microscopic localization of the neural recognition molecules L1, NCAM, and its isoform NCAM180, the NCAM-associated polysialic acid, beta1 integrin and the extracellular matrix molecule tenascin-R in synapses of the adult rat hippocampus. *Journal of neurobiology,* 49, 142-58.

SCHWARZ, V., PAN, J., VOLTMER-IRSCH, S. & HUTTER, H. 2009. IgCAMs redundantly control axon navigation in Caenorhabditis elegans. *Neural Dev,* 4, 13.

SEAL, R. L., GORDON, S. M., LUSH, M. J., WRIGHT, M. W. & BRUFORD, E. A. 2011. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res,* 39, D514-9.

SEBAT, J., LAKSHMI, B., TROGE, J., ALEXANDER, J., YOUNG, J., LUNDIN, P., MANER, S., MASSA, H., WALKER, M., CHI, M., NAVIN, N., LUCITO, R., HEALY, J., HICKS, J., YE, K., REINER, A.,

GILLIAM, T. C., TRASK, B., PATTERSON, N., ZETTERBERG, A. & WIGLER, M. 2004. Large-scale copy number polymorphism in the human genome. *Science,* 305, 525-8.
SELLAR, G. C., WATT, K. P., RABIASZ, G. J., STRONACH, E. A., LI, L., MILLER, E. P., MASSIE, C. E., MILLER, J., CONTRERAS-MOREIRA, B., SCOTT, D., BROWN, I., WILLIAMS, A. R., BATES, P. A., SMYTH, J. F. & GABRA, H. 2003. OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. *Nat Genet,* 34, 337-43.
SHAPIRO, L., LOVE, J. & COLMAN, D. R. 2007. Adhesion Molecules in the Nervous System: Structural Insights into Function and Diversity. *Annual Review of Neuroscience,* 30, 451.
SHERIDAN, J. D. 1968. Electrophysiological evidence for low-resistance intercellular junctions in the early chick embryo. *The Journal of cell biology,* 37, 650-9.
SHIMOYAMA, M., SMITH, J. R., HAYMAN, T., LAULEDERKIND, S., LOWRY, T., NIGAM, R., PETRI, V., WANG, S. J., DWINELL, M., JACOB, H. & TEAM, R. G. D. 2011. RGD: a comparative genomics platform. *Hum Genomics,* 5, 124-9.
SIEGEL, P. B., DODGSON, J. B. & ANDERSSON, L. 2006. Progress from chicken genetics to the chicken genome. *Poult Sci,* 85, 2050-60.
SINGER, M. F. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell,* 28, 433-4.
SMITH, J. L. & SCHOENWOLF, G. C. 1988. Role of cell-cycle in regulating neuroepithelial cell shape during bending of the chick neural plate. *Cell Tissue Res,* 252, 491-500.
SPALTMANN, F. & BRUMMENDORF, T. 1996. CEPU-1, a novel immunoglobulin superfamily molecule, is expressed by developing cerebellar Purkinje cells. *J Neurosci,* 16, 1770-9.
SPILIANAKIS, C. G., LALIOTI, M. D., TOWN, T., LEE, G. R. & FLAVELL, R. A. 2005. Interchromosomal associations between alternatively expressed loci. *Nature,* 435, 637-45.
STAUNTON, D. E., DUSTIN, M. L., ERICKSON, H. P. & SPRINGER, T. A. 1990. The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell,* 61, 243-54.
STRUYK, A. F., CANOLL, P. D., WOLFGANG, M. J., ROSEN, C. L., D'EUSTACHIO, P. & SALZER, J. L. 1995. Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. *J Neurosci,* 15, 2141-56.
STULEN, G. 1981. Electric field effects on lipid membrane structure. *Biochim Biophys Acta,* 640, 621-7.
SUGIMOTO, C., MAEKAWA, S. & MIYATA, S. 2010. OBCAM, an immunoglobulin superfamily cell adhesion molecule, regulates morphology and proliferation of cerebral astrocytes. *Journal of neurochemistry,* 112, 818-28.
TAM, P. P. L. & ROSSANT, J. 2003. Mouse embryonic chimeras: tools for studying mammalian development. *Development,* 130, 6155-6163.
TANABE, Y., WILLIAM, C. & JESSELL, T. M. 1998. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell,* 95, 67-80.

TEISSIE, J. & ROLS, M. P. 1993. An experimental evaluation of the critical potential difference inducing cell membrane electropermeabilization. *Biophys J,* 65, 409-13.
TESSIER-LAVIGNE, M. & GOODMAN, C. S. 1996. The molecular biology of axon guidance. *Science,* 274, 1123-33.
THOMAS, K. R. & CAPECCHI, M. R. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell,* 51, 503-512.
TIMMER, J., JOHNSON, J. & NISWANDER, L. 2001. The use of in ovo electroporation for the rapid analysis of neural-specific murine enhancers. *Genesis,* 29, 123-32.
TREIT, D., ENGIN, E. & MCEOWN, K. 2010. Animal models of anxiety and anxiolytic drug action. *Current topics in behavioral neurosciences,* 2, 121-60.
TURAN, S. & BODE, J. 2011. Site-specific recombinases: from tag-and-target-to tag-and-exchange-based genomic modifications. *FASEB J,* 25, 4088-107.
TURNPENNY, P. D. & ELLARD, S. 2005. *Emery's elements of medical genetics,* Edinburgh ; New York, Elsevier/Churchill Livingstone.
TURNPENNY, P. D., ELLARD, S. & SCIENCEDIRECT (ONLINE SERVICE) 2012. Emery's elements of medical genetics. 14th ed. Philadelphia, PA: Elsevier/Churchill Livingstone.
UNTERGASSER, A., NIJVEEN, H., RAO, X., BISSELING, T., GEURTS, R. & LEUNISSEN, J. A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res,* 35, W71-4.
VAN DE LAVOIR, M. C., DIAMOND, J. H., LEIGHTON, P. A., MATHER-LOVE, C., HEYER, B. S., BRADSHAW, R., KERCHNER, A., HOOI, L. T., GESSARO, T. M., SWANBERG, S. E., DELANY, M. E. & ETCHES, R. J. 2006. Germline transmission of genetically modified primordial germ cells. *Nature,* 441, 766-9.
VAN GENDEREN, C., OKAMURA, R. M., FARIÑAS, I., QUO, R. G., PARSLOW, T. G., BRUHN, L. & GROSSCHEDL, R. 1994. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes & Development,* 8, 2691-2703.
VESTAL, D. J. & RANSCHT, B. 1992. Glycosyl phosphatidylinositol--anchored T-cadherin mediates calcium-dependent, homophilic cell adhesion. *The Journal of cell biology,* 119, 451-61.
VESTWEBER, D. 2000. Molecular mechanisms that control endothelial cell contacts. *Journal of Pathology,* 190, 281-291.
VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457, 854-8.
VISEL, A., MINOVITSKY, S., DUBCHAK, I. & PENNACCHIO, L. A. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research,* 35, D88-92.
VO, N. & GOODMAN, R. H. 2001. CREB-binding protein and p300 in transcriptional regulation. *The Journal of biological chemistry,* 276, 13505-8.

VOLKMER, H., SCHREIBER, J. & RATHJEN, F. G. 2012. Regulation of Adhesion by Flexible Ectodomains of IgCAMs. *Neurochem Res.*

WALL, L., DEBOER, E. & GROSVELD, F. 1988. The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes & development,* 2, 1089-100.

WALMOD, P. S., PEDERSEN, M. V., BEREZIN, V. & BOCK, E. 2007. Cell Adhesion Molecules of the Immunoglobulin Superfamily in the Nervous System. *In:* LAJTHA, A., BANIK, N. L. & BANIK, N. (eds.) *Handbook of Neurochemistry and Molecular Neurobiology: Neural protein metabolism and function.* Berlin: Springer-Verlag.

WALSH, F. S. & DOHERTY, P. 1997. Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. *Annu Rev Cell Dev Biol,* 13, 425-56.

WANG, L., HAUSER, E. R., SHAH, S. H., SEO, D., SIVASHANMUGAM, P., EXUM, S. T., GREGORY, S. G., GRANGER, C. B., HAINES, J. L., JONES, C. J., CROSSMAN, D., HAYNES, C., KRAUS, W. E., FREEDMAN, N. J., PERICAK-VANCE, M. A., GOLDSCHMIDT-CLERMONT, P. J. & VANCE, J. M. 2008. Polymorphisms of the tumor suppressor gene LSAMP are associated with left main coronary artery disease. *Ann Hum Genet,* 72, 443-53.

WANG, X. & EL NAQA, I. M. 2008. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics,* 24, 325-32.

WASSARMAN, P. M. & SORIANO, P. M. 2010. *Guide to techniques in mouse development,* San Diego, Calif. ; London, Elsevier/Academic Press.

WEAVER, J. C. 1995. Electroporation theory. Concepts and mechanisms. *Methods Mol Biol,* 55, 3-28.

WICHTERLE, H., LIEBERAM, I., PORTER, J. A. & JESSELL, T. M. 2002. Directed differentiation of embryonic stem cells into motor neurons. *Cell,* 110, 385-97.

WILKINSON, P., SENGEROVA, J., MATTEONI, R., CHEN, C.-K., SOULAT, G., URETA-VIDAL, A., FESSELE, S., HAGN, M., MASSIMI, M., PICKFORD, K., BUTLER, R. H., MARSCHALL, S., MALLON, A.-M., PICKARD, A., RASPA, M., SCAVIZZI, F., FRAY, M., LARRIGALDIE, V., LEYRITZ, J., BIRNEY, E., TOCCHINI-VALENTINI, G. P., BROWN, S., HERAULT, Y., MONTOLIU, L., DE ANGELIS, M. H. & SMEDLEY, D. 2010. EMMA—mouse mutant resources for the international scientific community. *Nucleic Acids Research,* 38, D570-D576.

WILLIAMS, A. F. & BARCLAY, A. N. 1988. The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol,* 6, 381-405.

WILSON, D. B. 1973. Chronological changes in the cell cycle of chick neuroepithelial cells. *Journal of embryology and experimental morphology,* 29, 745-51.

WILSON, D. J., KIM, D. S., CLARKE, G. A., MARSHALL-CLARKE, S. & MOSS, D. J. 1996. A family of glycoproteins (GP55), which inhibit neurite outgrowth, are members of the Ig superfamily and are related to OBCAM, neurotrimin, LAMP and CEPU-1. *J Cell Sci,* 109 (Pt 13), 3129-38.

WOESE, C. R., KANDLER, O. & WHEELIS, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A,* 87, 4576-9.

WOOLFE, A., GOODSON, M., GOODE, D. K., SNELL, P., MCEWEN, G. K., VAVOURI, T., SMITH, S. F., NORTH, P., CALLAWAY, H., KELLY, K., WALTER, K., ABNIZOVA, I., GILKS, W., EDWARDS, Y. J., COOKE, J. E. & ELGAR, G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol,* 3, e7.

WORLD HEALTH ORGANISATION. 2006. *Sickle-cell anaemia : report by the Secretariat,* Geneva, World Health Organization.

WU, C., OROZCO, C., BOYER, J., LEGLISE, M., GOODALE, J., BATALOV, S., HODGE, C. L., HAASE, J., JANES, J., HUSS, J. W., 3RD & SU, A. I. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol,* 10, R130.

WU, L. C., SUN, C. W., RYAN, T. M., PAWLIK, K. M., REN, J. & TOWNES, T. M. 2006. Correction of sickle cell disease by homologous recombination in embryonic stem cells. *Blood,* 108, 1183-8.

XENARIOS, I., RICE, D. W., SALWINSKI, L., BARON, M. K., MARCOTTE, E. M. & EISENBERG, D. 2000. DIP: the database of interacting proteins. *Nucleic acids research,* 28, 289-91.

XIE, Y., SUN, T., WANG, Q. T., WANG, Y., WANG, F., PUSCHECK, E. & RAPPOLEE, D. A. 2005. Acquisition of essential somatic cell cycle regulatory protein expression and implied activity occurs at the second to third cell division in mouse preimplantation embryos. *FEBS Lett,* 579, 398-408.

YAMADA, M., HASHIMOTO, T., HAYASHI, N., HIGUCHI, M., MURAKAMI, A., NAKASHIMA, T., MAEKAWA, S. & MIYATA, S. 2007. Synaptic adhesion molecule OBCAM; synaptogenesis and dynamic internalization. *Brain Res,* 1165, 5-14.

YAMAGATA, M., WEINER, J. A. & SANES, J. R. 2002. Sidekicks: Synaptic Adhesion Molecules that Promote Lamina-Specific Connectivity in the Retina. *Cell,* 110, 649-660.

YANEZA, M., GILTHORPE, J. D., LUMSDEN, A. & TUCKER, A. S. 2002. No evidence for ventrally migrating neural tube cells from the mid- and hindbrain. *Dev Dyn,* 223, 163-7.

YEE, S. P. & RIGBY, P. W. 1993. The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev,* 7, 1277-89.

YEN, C. C., CHEN, W. M., CHEN, T. H., CHEN, W. Y., CHEN, P. C., CHIOU, H. J., HUNG, G. Y., WU, H. T., WEI, C. J., SHIAU, C. Y., WU, Y. C., CHAO, T. C., TZENG, C. H., CHEN, P. M., LIN, C. H., CHEN, Y. J. & FLETCHER, J. A. 2009. Identification of chromosomal aberrations associated with disease progression and a novel 3q13.31 deletion involving LSAMP gene in osteosarcoma. *Int J Oncol,* 35, 775-88.

ZACCO, A., COOPER, V., CHANTLER, P. D., FISHER-HYLAND, S., HORTON, H. L. & LEVITT, P. 1990. Isolation, biochemical characterization and ultrastructural analysis of the limbic system-associated membrane protein (LAMP), a protein expressed by neurons comprising functional neural circuits. *J Neurosci,* 10, 73-90.

ZHANG, C., XUAN, Z., OTTO, S., HOVER, J. R., MCCORKLE, S. R., MANDEL, G. & ZHANG, M. Q. 2006. A clustering property of highly-

degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Research,* 34, 2238-46.

ZHU, X. D. & SADOWSKI, P. D. 1995. Cleavage-dependent ligation by the FLP recombinase. Characterization of a mutant FLP protein with an alteration in a catalytic amino acid. *J Biol Chem,* 270, 23044-54.

ZHUKAREVA, V. & LEVITT, P. 1995. The limbic system-associated membrane protein (LAMP) selectively mediates interactions with specific central neuron populations. *Development,* 121, 1161-72.

ZIV, N. E. & SMITH, S. J. 1996. Evidence for a role of dendritic filopodia in synaptogenesis and spine formation. *Neuron,* 17, 91-102.